Methods for Bibliometric Analysis of Research: Renewable Energy Case Study

Blaine Ziegler

Working Paper CISL# 2009-10

September 2009

Composite Information Systems Laboratory (CISL)
Sloan School of Management, Room E53-320
Massachusetts Institute of Technology
Cambridge, MA 02142

Methods for Bibliometric Analysis of Research: Renewable Energy Case Study

by

Blaine E. Ziegler

S.B., E.E.C.S. M.I.T., 2008

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

August 2009

Author_____
Department of Electrical Engineering and Computer Science
August 21, 2009

Certified by_____
Stuart Madnick
Professor of Information Technology
Thesis Supervisor

Accepted by_____
Arthur C. Smith
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Theses

Methods for Bibliometric Analysis of Research: Renewable Energy Case Study
by
Blaine E. Ziegler

Submitted to the
Department of Electrical Engineering and Computer Science

August 21, 2009

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

This paper presents methods and software implementation for analyzing a field of research through the use of bibliometrics, i.e., information about published journal articles. Online publication search engines are queried, and their search results are extracted and analyzed, to help inform a researcher of the state of his or her field. Our methodology consists of three components: extraction of terms relevant to the research field, analysis of the growth in prevalence of these terms over time, and identification of interrelationships among these terms using a technique known as Latent Semantic Analysis. These methods are applicable to the analysis of any research field, but this paper presents results from a case study on the field of renewable energy.

Thesis Supervisor: Stuart Madnick
Title: Professor of Information Technology

Contents

Chapter 1: Overview and Objectives

1.1 Introduction

For researchers in a technical field, understanding the state of their area of interest is of the highest importance. Any research field is composed of many subfields and underlying technologies which are related in intricate ways. This composition, or "research landscape," is not static. New technologies are constantly developed, while old ones become obsolete. Fields that are presently unrelated may one day become dependent on each other's findings. An invention from decades prior may find a new application in an emerging field.

Experts in a field could already have a strong understanding of their research landscape. But it would be unreasonable to expect them to have intimate knowledge of every aspect of their field. More critically, fledgling technologies that could one day play an important role are unfortunately the ones of which they are most likely unaware. A researcher may attempt to remain up to date by consulting news services or blogs, or by conversing with other experts. But these activities are time consuming and imprecise, and they do not ensure that the researcher will not miss an important development, particularly in subfields on which the researcher is not an expert.

The field of library and information science is concerned with the classification and organization of information. Bibliometric analysis, the extraction of statistics on journal articles and the research fields they discuss, comprises some of the tools used in library science. Bibliometric analysis (or bibliometrics) may be used, for example, to compare the prevalence of multiple fields of research. Bibliometrics can also be used to determine the relevance of a particular article to a particular field by considering the citations listed in each of a set of articles.

The field of data mining deals with the challenge of extracting useable information from large and complex data sets. The use of data mining when applied to the bibliometric analysis of technology fields is known as tech mining. In this case, research fields are the variables among which we are attempting to uncover relationships. We may discover, for example, that articles discussing biomass tend to also often discuss renewable energy, from which we may conclude that biomass and renewable energy are related fields of research. Tech mining can also be used to identify the rate of change of the prevalence of a particular technology. This could be useful, for example, in identifying R&D projects in which to invest.

In this project, we apply novel approaches, as well as established ideas from the technology forecasting and bibliometrics literature, to help meet the needs of researchers in terms of fully understanding the relationships between their field and others, and the change in relative prevalence of these fields over time. We propose automated methods for mapping the current research landscape and extrapolating the increase in prevalence of emerging technologies over time.

Bibliometric information is extracted from online publication search engines such as Engineering Village[1] and Scirus[2], and this information is analyzed for insight on a research field of interest. In particular, two types of information from these online databases were found to be especially useful: the number of documents related to a given keyword (the "hit count"), and a set of other terms related to the given keyword. (The phrases "keyword" and "term" are used interchangeably throughout this paper.) Using this information, we perform various analyses for research landscape mapping and technology forecasting.

Although the methods proposed in this paper should apply equally well to any field of research, we focus our discussion with a case study regarding renewable energy technologies. This field was chosen because it encompasses many highly active subfields and is an example of diverse research fields being used in conjunction for new purposes; e.g., biofuel development depends in part on genetic engineering, and efficient wind power generation requires high quality lubricants. Throughout this paper, concrete results will be presented for the renewable energy field.

1.2 Objectives

The aim of this project is to use computational and algorithmic techniques to automatically extract and analyze information from online publication search engines, for the purpose of identifying important descriptors of the state of a field of research. Specifically, we build a suite of software that achieves the following objectives:

- **Discovers related technologies**

Our software takes as input a particular field of interest (renewable energy, for example) and uses information available from online publication search engines to create a list of related terms or technologies. The length of the resultant list is an adjustable parameter, but we typically aim for a few thousand items. The utility of this list is twofold; first, it may contain technologies or research areas that interest an expert in the field, and second, the list is used as the input to other pieces of software for further processing and analysis.

- **Calculates a numerical value for the growth rate of a technology**

Online publication search engines are used to find the number of articles published in a technological field in each year over a given time span, which is typically five to ten years. These yearly publication counts are then analyzed to calculate a numerical index of the technology's rate of growth. This is useful for determining to what extent interest in a technology is rising or falling. A range of such indices are screened for applicability as indicators of potential.

- **Determines the relationships among a set of technologies**

[1] *www.engineeringvillage.com*
[2] *www.scirus.com*

Given a set of technologies, techniques are proposed for organizing them in logical groupings. The information for performing this calculation comes from co-occurrences of different technologies across the documents indexed by an online publication search engine. For example, if many documents contain both terms *solar cell* and *solar power*, it may be concluded that these two technologies are related, and should hence be grouped together. This is useful for describing the research landscape of a field.

- **Identifies important, fast-growing technologies in which an expert may be interested**

The techniques described above are combined to create a system which takes as input a single technology, produces a list of related technologies, ranks and filters them according to growth rate, and combines the remaining technologies into logical groups. The output is then a list of fast-growing technologies which are related to the input technology and which have been organized into self-similar groupings. It is believed that this information would be helpful to an expert for deciding which technologies are worth observing more closely in the near future.

- **Validates our methods of technology identification by testing on fast-growing technologies of the past**

We back-test our methods by considering technologies that had a fast growth in prevalence in the past, and determining whether our software would have identified those technologies as such, given only information available before its growth in popularity. The successful technologies on which our software is back-tested are those with a sufficiently large long-term growth rate.

1.3 Overview of methods

Later chapters discuss the methods used to achieve the above objectives in detail, but we present a general overview in this section.

Note: In this paper we abstract the concept of a research field into a single keyword. It is a keyword in the sense that it can be used to query a search engine and obtain information about the field. When we discuss a particular keyword (or term), it is understood to represent a technology or research field. For example, if we say that a search for the keyword *renewable energy* returns the related keywords *wind power* and *solar power*, this implies that the wind power and solar power fields of research are related to the field of renewable energy. We will make extensive use of this shorthand terminology of labeling a research field as a keyword or term.

1.3.1 Discovering related technologies

Given a particular keyword, related technologies can be identified by querying online publication search engines, using the keyword as the search term. The results that are returned from these search engines contain, in some form, a set of terms that are potentially relevant. Our software extracts these terms, after which various filtering techniques can be applied to derive a list of terms that are highly relevant to the given technology, as well as potentially interesting to an expert in the field.

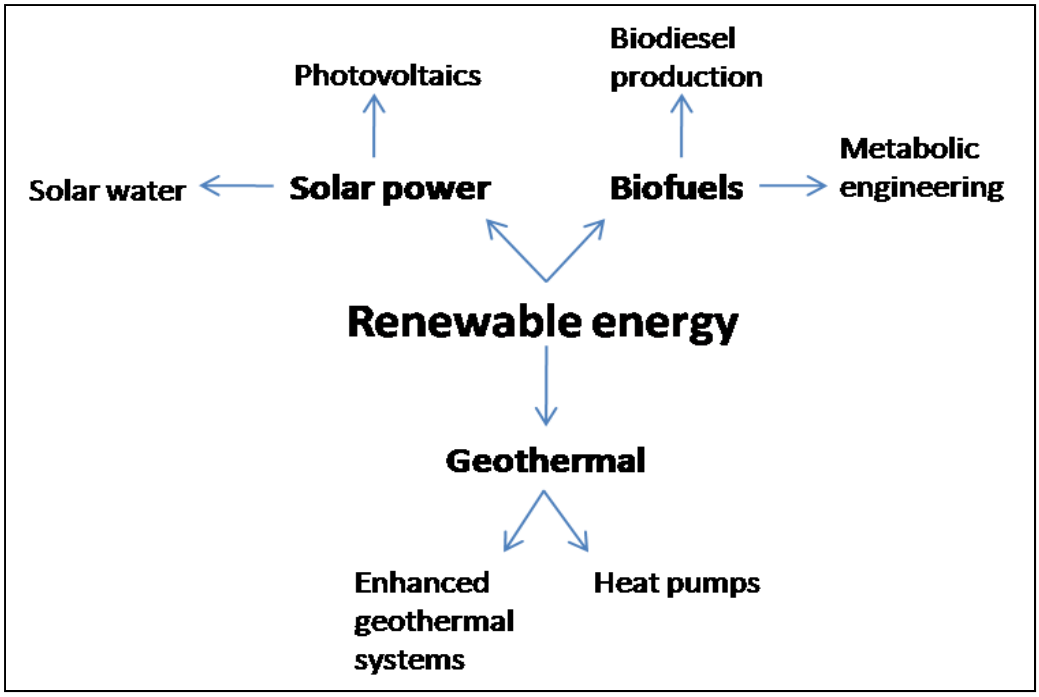

Figure 1. Hierarchical representation of renewable energy-related technologies.

We use two different search engines for the identification of related technologies: Engineering Village and Scirus. Both provide a means of extracting terms that are related to the given keyword, whereas many other search engines do not. Engineering Village and Scirus present their related keywords in two entirely different ways however, so each requires its own unique method of term extraction.

The search results for Engineering Village consist of a list of articles, sorted by relevance to the keyword. Associated with each article is a set of keywords that the article's author has used to characterize the article. Therefore, to use Engineering Village for the extraction of related technologies, our software queries the search engine with the given technology as a keyword, and extracts from the most relevant articles (as determined by the search engine) the associated author-created keywords.

The search results for Scirus also consist of a list of articles, but in this case, there is not a list of keywords associated with each article. However, what Scirus does provide is a "refine your search" feature consisting of a list of 80 keywords that it considers relevant to the given search term. In other words, keywords are provided that are relevant to the search term itself, rather than to the individual articles about the search term. Since a Scirus search for a given term only produces 80 related terms, which is fewer than we desire, we perform follow-up searches on these 80 terms to discover far more. The number of terms acquired from this second-level search is upper-bounded at 6,400, but the number of non-repeated terms is smaller. We can continue performing follow-up searches to extract terms from Scirus in a hierarchical manner. The breadth and depth of this hierarchy is an adjustable parameter that allows the extraction of as many terms as needed.

1.3.2 Calculating the growth rate of a technology

The rate at which a technology is growing can be determined by analyzing the trend of its hit counts over time. A range of years is chosen over which we are interested in the prevalence of the technology. Then for each year, a publication search engine is queried for the number of related articles which were published in that year. This provides a time series of the technology's hit counts, which can be used to calculate a numerical index of the growth rate of the technology's prevalence over time.

There are many options for the specific method for deriving a single growth rate index from a time series of hit counts. One simple but effective strategy is to calculate the ratio of the final year's hit count to the initial year's hit count, ignoring intermediate years. A similar strategy is to take the difference, rather than the ratio, of the final and initial hit counts. A third method fits an exponential curve to the time series and uses the best-fit exponent as the growth rate. Still another metric is the technology's average year of publication. These methods have their own unique strengths and weaknesses. But whatever definition of growth rate we choose, we can use it to rank a set of technologies and demonstrate which are growing quickly and which are stagnating or declining.

1.3.3 Determining the relationships among technologies

We use a mathematical technique known as Latent Semantic Analysis (LSA) to reveal relationships among a set of technologies. Each technology is represented as a keyword which is submitted to a publication search engine. LSA analyzes the co-occurrences of each of the terms in the documents indexed by the search engine to identify groupings of related terms.

LSA groups individual terms together into the concepts that underlie them. For example, consider the terms *storm*, *lightning*, *bolt*, *nut*, *muffin*, and *whale*. Intuitively, these terms represent four distinct underlying concepts. *Lightning* and *bolt* are part of a concept about storms, *bolt* and *nut* are part of a concept about hardware, *nut* and *muffin* make up a concept about pastries, and *whale* represents its own concept. Note that a single term may be part of multiple concepts.

By searching the documents indexed by a search engine, we could determine that *lightning* and *bolt* frequently co-occur, as do *bolt*/*nut* and *nut*/*muffin*, but that *bolt* and *muffin*, for example, do not. LSA uses this information to find and return the four underlying concepts, each of which is a grouping of the terms of which it is composed.

For the purpose of identifying the relationships among technologies, the underlying concepts are less obvious. They may represent fields of research or products, which are composed of a set of specific technologies. And as in the example above with *bolt* and *nut*, some specific technologies may be part of multiple fields. LSA provides a promising method for identifying these relationships.

1.3.4 Identifying technologies interesting to an expert

Used in combination, the software components described thus far can help to identify emerging technologies that an expert might find useful. First, the term collection software is used to identify relevant technologies; next hit count analysis is used to reduce the list to only those technologies that

are growing at a high rate. Finally, LSA organizes this list into groups of related technologies so that we can observe the growth of fields that rely on multiple technologies. The output of this entire process – a list of technologies and fields that are growing quickly and related to a given field of interest – can be presented to an expert for review.

1.3.5 Validating our methods

One of the most important goals of this project is to automatically identify new, fast-growing technologies. We can assess the usefulness of our methods by back-testing; i.e. determining whether our software would have identified the emergence of such technologies, given only the information available before their rise in prevalence. Given a list of potential terms for back-testing, we consider only those whose long-term growth rates (over approximately ten years) meet some minimum threshold. Our system is presented with the information that was available during the very early stages of the technology's growth. Its usefulness is then evaluated in terms of its success in predicting technologies that ultimately experience large growths in prevalence.

1.4 Thesis overview

This chapter has presented the objectives of our project, along with a brief description of the methods used to achieve these objectives. Chapter 2, *Technical Introduction and Literature Review*, presents a review of the research and literature in the fields of technology forecasting, bibliometrics, and tech mining. Chapter 3, *Keyword Extraction and Analysis*, describes how we extract keywords from online publication search engines, how the associated publication counts are extracted, and how we use these counts to identify fast-growing technologies. Chapter 4, *Latent Semantic Analysis*, describes the well-known LSA algorithm, which is used to group related keywords into higher-level concepts. Our novel modifications and data cleaning methods, which help to adapt the algorithm meet our needs, are also discussed. Chapter 5, *Renewable Energy Case Study Results*, presents results of an experiment in which we combine the pieces of our software to analyze the field of renewable energy. Finally, Chapter 6, *Conclusions*, summarizes the main findings of the project and, in particular, discusses how the objectives listed in Chapter 1 were met.

Chapter 2: Literature Review

The work presented in this paper is a form of technology forecasting. Although we stop short of claiming to predict the future, our methodology includes technology growth rate analysis, which is a suggestion of future prevalence. Our methods are derived from the field of bibliometric analysis. The first section of this chapter discusses the literature surrounding the field of technology forecasting, and the second section discusses bibliometrics.

2.1 Technology forecasting

Technology forecasting is the process of forming predictions about the future state of a technology. One may be interested in its future prevalence in terms of consumer uptake, research activity, profitability, etc. Numerous techniques exist for this activity. Some are well-defined and studied methods, while others are less formal. A brief review is presented here, and the reader is referred to [Porter et al., 2004] and [Kaya Firat et al., 2008] for more extensive surveys of the field.

One of the most widely-practiced methodologies of technology forecasting is consideration of expert opinion. Subject area experts are asked to make predictions on the direction in which their field is heading. The fact that this is a form of technology forecasting is often implicit; any organization is always gathering expert opinion for insight on the technologies to which it should devote its resources. However, more structured methods also exist, with the explicit purpose of technology forecasting via expert opinion. One commonly used technique is the Delphi method. In a panel of experts, each is given a survey with questions involving his or her opinion on the future of a technology. The surveys are answered anonymously to encourage honest opinions. All surveys are read by a facilitator, and the key points from the group are compiled and recited to the panel as a whole. After having heard the opinions of the group, all participants are asked to again answer the survey questions, revising their previous responses if they wish. This process may continue for several rounds. In theory, the group eventually arrives at a consensus, and this group-wide forecast is thought to be more reliable than any individual opinion from the first round of responses [Rowe and Wright, 2001].

Two techniques that can assist in expert forecasting are the so-called “futures wheel” and “relevance tree.” These techniques break down a forecast into smaller, more manageable pieces. A futures wheel is a diagram which at its center contains an event or trend surrounding the technology which may have an impact on its future. This could be, for example, a recent change in research funding, or a trend towards different consumer habits. The likely effects of this event are then added to the futures wheel in nodes immediately surrounding the center. In turn, the likely effects of these primary effects are added to produce another ring of the wheel. After several iterations of this process, the wheel begins to represent a forecast, with events listed on the outside edge theorized to occur in the future [Glenn, 1972]. The relevance tree can be thought of as the opposite of a futures wheel. Whereas a futures wheel begins with a known event and branches out to potential consequences, a relevance tree begins with a possible future scenario, such as the uptake of a technology, and hierarchically decomposes this goal into the objectives that must first be achieved for this to occur. For each objective or sub-objective in the tree, conditional probabilities are estimated for the occurrence of

that objective, given the occurrence of the sub-objective below it. Using these conditional probabilities, an estimated probability can be calculated for the main objective at the root of the tree [Levary and Han, 1995].

Environmental scanning is another method of technology forecasting. Scanning involves observing current events and indicators to become better informed on the direction in which a technology may be headed. According to [Aguilar, 1967], the purpose of disciplined scanning is to structure the information considered by an organization and to alert it when key indicators change. [Morrison et al., 1984] defined a distinction between passive scanning and active scanning. Passive scanning is the type of implicit observation that individuals and organizations do on a regular basis, such as watching the news, etc. Active scanning, on the other hand, involves regular review of specific field-related resources. The environments scanned in an active approach include the *task environment*, *industry environment*, and *macro environment*. Defined by [Fahey and Narayanan, 1986], the task environment refers to an organization's set of customers, the industry environment to the set of related institutions, and the macro environment to the "social, technological, economic, environmental, and political" conditions of the world. Any number of resources may be used for scanning, but common ones include newspapers, journals, political literature, Census Bureau data, and subject area experts [Morrison, 1992].

Another technology forecasting technique is scenario analysis. This involves envisioning various possible states of a future world. In each state, the technology of interest has a particular degree of prevalence. The likelihood of each state is estimated, and from this, the expected future prevalence of the technology can be estimated as well [Levary and Han, 1995]. However, it is warned by [Schoemaker, 1998] that the effectiveness of scenario analysis can be limited somewhat by the biases or narrow imaginations of those involved.

Finally, we discuss trend analysis as a method for technology forecasting. Trend analysis involves the extrapolation of time series data in order to predict the future state of a technology. Extrapolation requires fitting a mathematical model to the data and using it to predict future data. It is often the case that the prevalence of a technology over time can be modeled by a logistic function, characterized by slow growth initially, followed by accelerated growth, and ultimately slowing again as its prevalence approaches a limit. The specific data used in trend analysis can vary, but a common metric is market share [Cuhls and Sahlo, 2003].

A component of our project involves trend analysis, using academic publication counts as the metric. It is in this sense that the project incorporates elements of both technology forecasting and bibliometric analysis.

2.2 Bibliometric analysis

Bibliometrics is a set of tools for analyzing publication data. [Norton, 2001] defines it as the measure of texts and information. The bibliometric information associated with a publication includes author, affiliation, citations from other publications, co-citations with other publications, reader usage, and associated keywords.

According to [Polanco, 1995], bibliometrics is a method for description, evaluation, and monitoring of research. It can describe the research surrounding a particular field, or similarly, it can describe the quantity and focus of research output by a particular organization. As an evaluation method, it can help determine the impact of a technology or the effectiveness of an author or research organization. Finally, it serves as a monitoring tool in that it can be used to track the level of activity in a research field over time.

[Porter, 2005] discusses so-called "tech mining," the processing of text databases to extract meaningful information on technologies of interest. As one example of tech mining, he presents techniques for identifying and visualizing keyword interrelationships. Identifying these relationships requires a metric for term similarity; [Saka and Igami, 2007] and [Small, 2006] utilize co-citation information for this purpose, and [Anuradha et al., 2007] and [Zhu and Porter, 2002] use author collaborations.

[Woon and Madnick, 2008] identify term relationships using co-occurrence within publications. This, along with the existence of powerful publication search engines, allows for utilization of vast databases, since all necessary information is represented by immediately accessible search engine hit counts. Specifically, the metric for term similarity (dissimilarity, rather) is known as the Normalized Google Distance, discussed in [Cilibrasi and Vitanyi, 2006]. The Normalized Google Distance defines the dissimilarity between two terms as a function of four quantities: the hit counts of each of the two terms, their hit count when queried jointly, and the total number of articles indexed by the search engine. Woon and Madnick use this metric to organize related terms into disjoint clusters.

An entirely different approach to organizing related terms is Latent Semantic Analysis. This technique is discussed in depth in Chapter 4 of this paper. But briefly, LSA uses occurrence counts of keywords in each of a body of documents to reveal the concepts underlying the set of terms. The concepts are identified using the mathematical technique known as Principal Component Analysis. LSA concepts differ from clusters, discussed above, because concepts are not disjoint; instead each term is present, to some degree, in each concept [Berry et al., 1995].

2.3 Novelty

This project makes use of hit count trend analysis, borrowing ideas from technology forecasting and bibliometrics. The novelty of our approach lies in the fact that the terms we analyze are generated in an automated manner. Experts are not asked to produce a list of terms, from which we identify the fastest-growing. Instead, our term extraction methodology generates the list with no human interaction. This is advantageous because it allows for the analysis of terms of which an expert may not have been aware.

Our project also makes use of LSA, an established technique. The novelty here is twofold. First, we make use of data cleaning techniques that detect and eliminate terms that are disruptive to the LSA results. Second, and perhaps more importantly, we modify the algorithm so that direct access to the text of all documents in a database is not required. Instead, we only require term hit counts, which are

readily available from publication search engines. This allows us to make use of the information contained in millions of publications.

Chapter 3: Keyword Extraction and Analysis

3.1 Overview

In seeking to gather information on the state of a field of research, the first step is the identification of a set of related subfields within the domain of interest. Analysis is then performed to infer how quickly each of these subfields is growing in prevalence. Our software queries online publication search engines to gather the necessary data, which are then investigated to draw conclusions about the field of interest. Specifically, the data we collect and analyze consist of sets of keywords and the associated publication counts of these terms over time.

This chapter describes the three stages of our keyword extraction and analysis methodology. The first piece, extraction, is described in section 3.2. In this stage, publication search engines are queried with a particular keyword, and related terms are extracted from the search results. This list of terms can be useful in its own right but is also put through further processing in later stages. The second piece, hit count extraction, is described in section 3.3. Here, publication search engines are queried for the publication counts (hit counts) over time of every keyword in our list. This produces a time series of the relative prevalence of the various technologies in our list. Finally, the third step, growth rate analysis, is described in section 3.4. In this piece, the hit count time series information collected in the previous step is used to calculate the growth rate of each term. The growth rate can be defined in one of several ways, with each having its own merits. By combining the three steps described in this chapter, we can provide as input a single technology, and receive as output a list of related technologies, ranked by rate of growth. This information can be useful to an expert interested in knowing about newly rising technologies in his or her field of research.

3.2 Keyword extraction

The first stage of analysis is keyword extraction. Given an initial seed term, we have developed software that automatically generates and submits queries to a range of publication search engines. The search engines respond with a results page containing, in one form or other, a listing of other, related keywords. These are located and extracted, then written to a file stored on the user's hard drive.

The choice of which search engines to use for keyword extraction is based on two important criteria: The first is that they must index a large number of articles related to the field of interest and spanning a range of publishers. Secondly, their search results must be formatted in such as way as to allow related keywords to be readable in an automated manner. The first criterion would exclude the use of small and publisher-specific databases such as IngentaConnect and SpringerLink. The second criterion excludes websites such as Google Scholar, which presents as its search results links to external websites, making the automated extraction of keywords difficult if not impossible.

Given these criteria, we have chosen two different publication search engines for keyword extraction, namely Engineering Village and Scirus. The two sites present related keywords in different

ways from one another, and as such require separate keyword extraction procedures. Our methods for extracting keywords from Engineering Village are described in section 3.2.1, and from Scirus in section 3.2.2.

3.2.1 Extraction from Engineering Village

This section describes our methodology for extracting keywords from the Engineering Village publication search engine. We first give an overview of Engineering Village and the way in which it presents related keywords. We then discuss how our software locates and extracts these keywords. Finally, we present results for keyword extraction with Engineering Village.

Engineering Village contains journal articles from three different publication databases: Compendex, Inspec, and NTIS. Users of Engineering Village can choose to return articles from any or all of these sources. For our purposes, our software only queries Compendex and Inspec from Engineering Village, and ignores NTIS. This is due to the fact that Compendex and Inspec contain many renewable energy-related articles, approximately one million each, whereas NTIS contains about half as many.

The format of the search results is the same regardless of whether the Compendex or Inspec databases are queried, as they are both accessible from within the Engineering Village website. As such, the procedure for keyword extraction is also the same. When queried with a search term a results page is returned which contains links to articles related to the topic, sorted in decreasing order of relevance. The precise manner in which Engineering Village defines relevance is unknown to us, but seems to correspond well with intuition. Twenty-five results are displayed per page, and a link to subsequent pages is provided.

Each link to an article in Engineering Village has a corresponding link to the article's abstract. When the user follows this link, a new page is displayed which contains the abstract as well as two sets of keywords that describe the article. The sets of keywords are labeled as "controlled" and "uncontrolled" terms. Controlled terms come from a list of pre-existing keywords that the publication database uses to characterize its articles. Every article's author is expected to choose a few (usually five to ten) of these keywords from the pre-compiled list to describe their paper. Uncontrolled terms are similar, with the difference being that the terms do not come from a pre-compiled list; instead the author is given the freedom to use any terms believed to best describe the article. As with the case of the controlled keywords, articles tend to have approximately five to ten uncontrolled keywords each.

For our purposes, both the controlled and uncontrolled terms can be useful. Controlled terms tend to be "safer" in that they are usually generally accepted terms to describe a field, and so the level of relevance to the original seed topic tends to be higher with controlled terms than with uncontrolled terms. On the other hand, the fact that uncontrolled terms are not necessarily universally known makes them useful for identifying emerging fields or technologies that are not yet well-established. Depending on the specific application, one or both sets of terms may be used.

The manner, then, in which keywords related to a given seed term are collected is to first query one of the databases, Compendex or Inspec, contained in Engineering Village. This returns a search

results page containing links to abstracts of articles, sorted by relevance. To obtain the corresponding keywords, the links for each of the first few articles in the list are followed. The number of articles that we consider is an adjustable parameter, but if the goal is to collect 1,000 terms, for example, approximately the first 100 articles would have to be considered. Each of these abstracts comes with a set of controlled terms and a set of uncontrolled terms which describe that particular article. We collect both sets of terms from each article with the presumption that they are relevant to our original seed term. Finally, we remove any duplicate terms that may exist, and the process is complete.

Screenshots are presented below that demonstrate the ability to extract keywords from Engineering Village.

Figure 2. Engineering Village search results page when queried with *renewable energy*. Provides links to abstracts, which contain keywords.

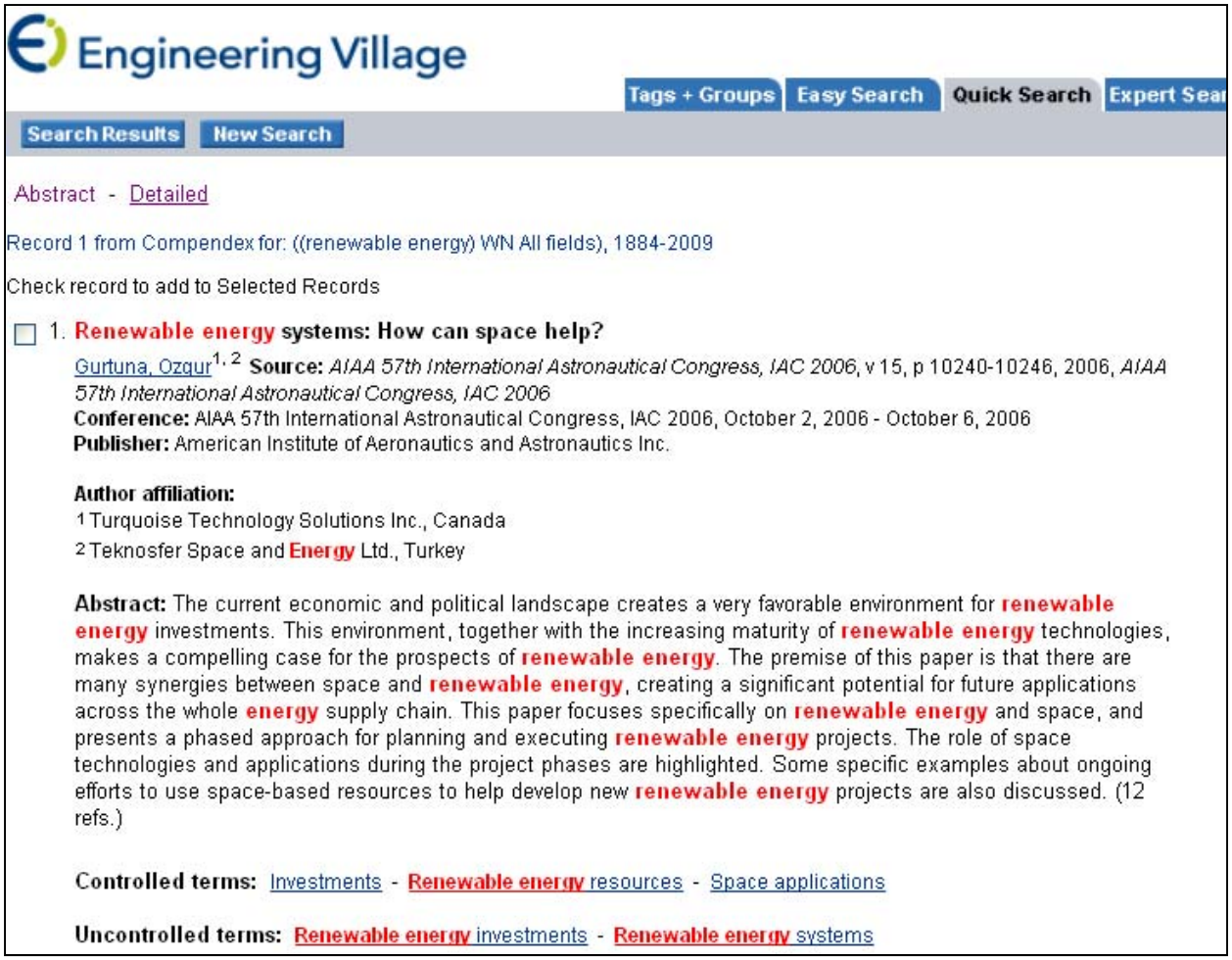

Figure 3. Engineering Village abstract page. Contains controlled and uncontrolled keywords beneath the abstract text.

3.2.1.1 Software implementation and performance

All of our software, including the Engineering Village term collection module, is written in the Python programming language. To execute this procedure, we call a function named *get_ev_terms*. This function takes three parameters: a list of keywords to use as seed terms, the number of documents from which to collect related terms, and an indication of whether we wish to use the Compendex or Inspec database.

While it is most common to use only a single term as a seed (*renewable energy,* for example*)*, collecting terms related to multiple keywords is also possible. In any case, for each seed term, the appropriate database is queried, and the returned results page is read into memory. This is done using Python's URL Library (*urllib*) module, which allows the contents of a web page to be read, given its URL. The URL of the desired search results page is constructed using a parameter string that replicates a standard query submitted using a web browser. The parameter string includes the search term as well as a value specifying whether Compendex or Inspec is to be searched.

Once the search results page has been read into memory, we use Python's Regular Expression module to extract specific pieces of this page. In particular, the page contains a list of document titles, along with a link to each document's abstract. Regular expressions are used to locate these links and extract the URLs to which they point.

Given a list of URLs to document abstracts, Python's *urllib* is again used to open each of these pages. Each page contains a list of both the controlled terms and uncontrolled terms that describe that particular document. Regular expressions are used to identify these lists of terms and read them into memory, maintaining separate lists for controlled terms and uncontrolled terms.

Finally, keywords must be read from the appropriate number of documents, as specified by the third parameter to the *get_ev_terms* function. Each search results page in Engineering Village contains links to the abstracts of 25 documents. If we wish to extract keywords from more than 25 documents, we must then navigate to successive search results pages. This is accomplished using regular expressions to identify the URL of the "Next page" link, and then using *urllib* to open that page.

Once we have populated our lists of controlled and uncontrolled terms from Compendex and/or Inspec, our software can merge these lists if we so choose. The merging process removes any duplicates (ignoring case), and alphabetizes the remaining terms. Finally, the resulting lists are written to disk for use in the following stages of the system.

The software for Engineering Village term collection runs in time proportional to the product of the number of seed terms provided and the number of documents from which we wish to collect the terms. The web page corresponding to each document must be accessed and its contents read into memory. This is the most significant performance bottleneck. The operation of using regular expressions to parse the page once it has been read requires negligible time.

As an example of the performance of this module, assuming that only a single seed term is used, it takes approximately 40 seconds to collect the terms from 100 documents, using a typical modern computer with a broadband internet connection. As we would expect, it takes approximately 80 seconds to collect terms from 200 documents, under the same conditions. (100 and 200 documents correspond to roughly 700 and 1,400 terms, respectively.) The largest source of variance in the runtime of this module is the speed of the connection to the online databases being used. It can run significantly slower if the local connection is slow or if the databases are experiencing heavy traffic.

3.2.1.2 Results

Results are now presented for the Engineering Village term extraction module, displaying terms related to *renewable energy*. Term lists are shown that were extracted from Compendex and Inspec, using both controlled and uncontrolled terms. In this section, all lists are derived from the ten most relevant articles. In appendix section A.1, we present larger sets of terms, using the fifty most relevant articles.

Compendex controlled terms, ten articles:

- *biomass*
- *building codes*
- *buildings*
- *carbon dioxide*
- *china*
- *computer aided instruction*
- *conservation*
- *construction industry*
- *curricula*
- *decision making*
- *design*
- *economic and social effects*
- *electric industry*
- *electric power plants*
- *energy conservation*
- *energy efficiency*
- *energy management*
- *energy policy*
- *energy utilization*
- *engineering education*
- *environmental impact*
- *gas emissions*
- *greenhouse effect*
- *greenhouse gases*
- *industrial economics*
- *intelligent buildings*
- *international law*
- *investments*
- *laws and legislation*
- *marketing*
- *mathematical models*
- *optimization*
- *personnel training*
- *pollution*
- *power generation*
- *regulatory compliance*
- *renewable energy resources*
- *sensitivity analysis*
- *social aspects*
- *societies and institutions*
- *space applications*
- *structural design*
- *sustainable development*
- *technology transfer*
- *world wide web*

Compendex uncontrolled terms, ten articles:

- *best practices*
- *building designers*
- *building regulations*
- *built environments*
- *calculation tools*
- *carbon cap-and-trade*
- *case studies*
- *development processes*
- *diffusion of renewables*
- *domestic buildings*
- *electricity sector*
- *energy conservation measures*
- *energy production*
- *energy security*
- *kirklees councils*
- *local authorities*
- *local government*
- *optimal renewable energy mathematical (orem) model*
- *policy makers*
- *policy options*
- *project datums*
- *refurbishment projects*
- *renewable energy*
- *renewable energy development*
- *renewable energy education*
- *renewable energy generation*
- *renewable energy investments*
- *renewable energy market*
- *renewable energy penetration*
- *renewable energy policy*
- *renewable energy projects*
- *renewable energy promotion law*
- *renewable energy statistics*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable portfolio standards*
- *renewables*
- *rural energy services*
- *sustainable buildings*
- *tertiary education*
- *testing phase*
- *wind facilities*

Inspec controlled terms, ten articles:

- *carbon compounds*
- *climatology*
- *commerce*
- *electricity supply industry*
- *energy conservation*
- *fossil fuels*
- *government policies*
- *legislation*
- *nuclear power stations*
- *power generation economics*
- *power utilisation*
- *reliability*
- *renewable energy sources*
- *sensitivity analysis*
- *socio-economic effects*
- *standardisation*
- *standards*
- *tariffs*

Inspec uncontrolled terms, ten articles:

- *accelerated depreciation plan*
- *agricultural sector*
- *atmospheric carbon dioxide*
- *baltic states*
- *biofuel*
- *biogas system*
- *biomass*
- *carbon dioxide emissions*
- *challenge 2008*
- *co2*
- *coal-fired power plants*
- *commercial sector*
- *competitive electricity supply market*
- *decarbonised world*
- *decentralised development*
- *delphi study*
- *developing countries*
- *domestic renewable energy generators*
- *earth climate change*
- *economic comparisons*
- *electricity prices*
- *electricity suppliers*
- *electricity supply industry*
- *energy conservation efforts*
- *energy demand variations*
- *energy diversification*
- *energy supply*
- *energy supply security considerations*
- *environmental friendly sustainable development*
- *estonia*
- *european union*
- *fossil fuel*
- *fossil fuel economy*
- *general public*
- *genoa*
- *geothermal park*
- *german renewable energy source act*
- *global renewable energy industry*
- *global renewable energy movement*
- *global warming*
- *government policy decisions*
- *government support*
- *green certificates*
- *green electricity*
- *greenhouse gas*
- *greenhouse gas reduction*
- *high cost*
- *human health*
- *hybrid development*
- *hydro power*
- *ignalina nuclear power plant*
- *india*
- *industrial sector*
- *institutional barriers*
- *international initiatives*
- *ipcc*
- *latvia*
- *legal barriers*
- *legal environment*
- *lighting*
- *lithuanian energy sector*
- *low energy density*
- *low-interest loans*
- *ministry of economic affairs of taiwan*
- *natural ecosystems*
- *nonpooled electricity trading*
- *oil*
- *optimal renewable energy mathematical model*
- *quota-based systems*
- *r&d*
- *reliability*
- *renewable energy company limited*
- *renewable energy development*
- *renewable energy development bill*
- *renewable energy development plan*
- *renewable energy development status*
- *renewable energy equipment procurement*
- *renewable energy policy harmonisation*
- *renewable energy programmes*
- *renewable energy promotion strategies*
- *renewable energy scenario*
- *renewable energy sources*
- *renewable energy systems allocation*
- *renewable energy systems cost*
- *renewable energy technologies*
- *renewable energy technology*
- *renewable energy trading experience*
- *renewable energy utilisation*
- *ring-fenced markets*
- *rural areas*
- *sensitivity analysis*
- *social acceptance variation*
- *socio economic optimal renewable energy model*
- *solar city*
- *solar energy*
- *solar pv*
- *statute for upgrading industries*
- *subsidy measures*
- *support mechanisms*
- *sustainable development*
- *sustainable energy source*
- *taiwan*
- *tax credit*
- *total energy demand*
- *transportation sector*
- *uk*
- *uk nffo*
- *urban areas*
- *waste energy resources*
- *wind energy*
- *wind farm*
- *wind power*

These results are promising in that they include many terms we would hope to find on such a list. We find, for example, *biomass*, *greenhouse gases*, *biofuel*, *biogas system*, *decarbonised world*, *geothermal park*, *waste energy resources*, *wind farm*, and many others that represent renewable energy technologies or policies. However, we also find terms whose level of generality makes them undesirable. For example, *mathematical models*, *optimization*, *best practices*, *case studies*, *sensitivity analysis*, *india*, and *general public* appear on these lists. While these terms are all somehow relevant to renewable energy, the connection may be loose or apply equally well to unrelated seed terms. Mathematical models and optimization, for example, are used in renewable energy research but are also used in unrelated areas. The terms are too widely applicable to be of much use.

We now briefly discuss the parameter choices for Engineering Village, with reference to the results shown here. There are two parameters: the choice of whether to use Compendex or Inspec, and the choice of whether to use controlled or uncontrolled keywords. There does not seem to be a clear difference between the terms extracted from Compendex and those extracted from Inspec. However, we note that Inspec seems to encourage a larger number of uncontrolled keywords per article than does Compendex. This has the potential of allowing for terms with greater specificity, since authors are not expected to make as much use of predefined terms when characterizing their articles.

The second parameter is the choice between controlled and uncontrolled keywords. Here, we see a relatively clear distinction in the results. Controlled terms tend to be general topics or areas of research; we see terms such as *conservation*, *decision making*, *design*, *electric industry*, *marketing*, *optimization*, *climatology*, and *legislation*. On the other hand, uncontrolled terms tend to be more specific. *Diffusion of renewables*, *kirklees councils*, *renewable energy promotion law*, *challenge 2008*, *competitive electricity supply market*, and *renewable energy systems allocation* are examples of this. As a general rule, an article's controlled terms tend describe the categories in which it belongs, while its uncontrolled terms describe what the article is actually about. The decision of which to use involves a tradeoff; controlled terms are usually relevant to the topic but may be too general, while uncontrolled terms are specific but can be irrelevant. For this project, in which the goal is to identify emerging technologies, we often choose in favor of specificity, with the risk of irrelevance. We therefore tend to prefer to use uncontrolled terms.

3.2.2 Extraction from Scirus

In this section, we discuss our methodology for using Scirus for keyword extraction. Due to the nature of the way in which search results are presented in Scirus, these methods are entirely different from those used for keyword extraction from Engineering Village. Also due to its unique presentation of search results, we have the opportunity to filter the terms we collect, by identifying and removing terms that are only loosely relevant to our seed topic. This section first describes basic keyword extraction, without filtering. We then present two different relevance filtering methods that we have devised.

3.2.2.1 Basic keyword extraction

Here we describe our basic approach to Scirus keyword extraction, without yet considering relevance filtering. We also discuss our software implementation and present results.

Scirus, the second search engine we use for keyword extraction, is different from Engineering Village in that its search results do not display journal articles in a consistent format within some Scirus template. Instead, its search results are links to outside websites which themselves contain articles in whatever format that particular site uses. For this reason, it is impossible to use Scirus for keyword collection in the same way that we used Engineering Village, i.e. visiting each article and collecting the keywords listed alongside the article's abstract.

Scirus does provide us with an entirely different way to collect related keywords, however. When a user searches Scirus for articles pertaining to a particular term, it returns a search results page that includes a "Refine your search" feature. This is a list of keywords related to the query, to aid the user in performing specific searches on their field of interest. The list always contains eighty terms, except in very rare cases when the seed term is obscure enough that fewer than eighty related terms are included.

So using Scirus, it is straightforward to extract eighty terms related to a given seed term. We simply query Scirus with our term, and collect the eighty terms that it lists as "refine your search" terms. It is important to reiterate that this is fundamentally different from the way terms are extracted from Engineering Village. With Engineering Village, we collect terms that describe journal articles that are relevant to the seed term. With Scirus, on the other hand, we collect terms that are directly related to the seed term, without considering articles at all.

While it is straightforward to extract eighty related keywords, we are usually interested in a much higher number of terms, generally on the order of a few thousand. This can be accomplished by collecting terms from Scirus in a hierarchical manner. First, Scirus is queried with the seed term, which gives eighty related terms. Subsequent queries are then conducted using any or all of these seed terms, resulting in a further eighty related terms for each of the original queries. This process may be iterated until we have collected as many keywords as desired.

The screenshots below demonstrate how keywords can be extracted from Scirus.

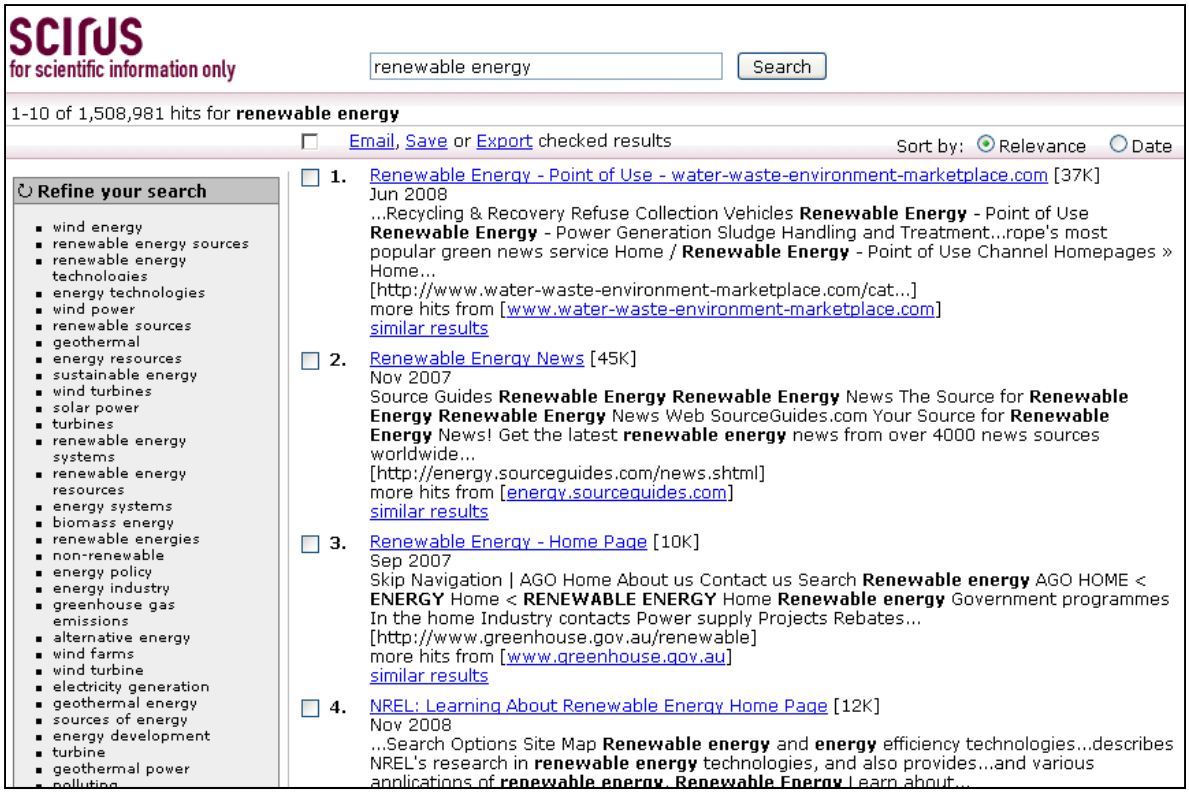

Figure 4. A query of *renewable energy* results in a list of related keywords in the "Refine your search" panel.

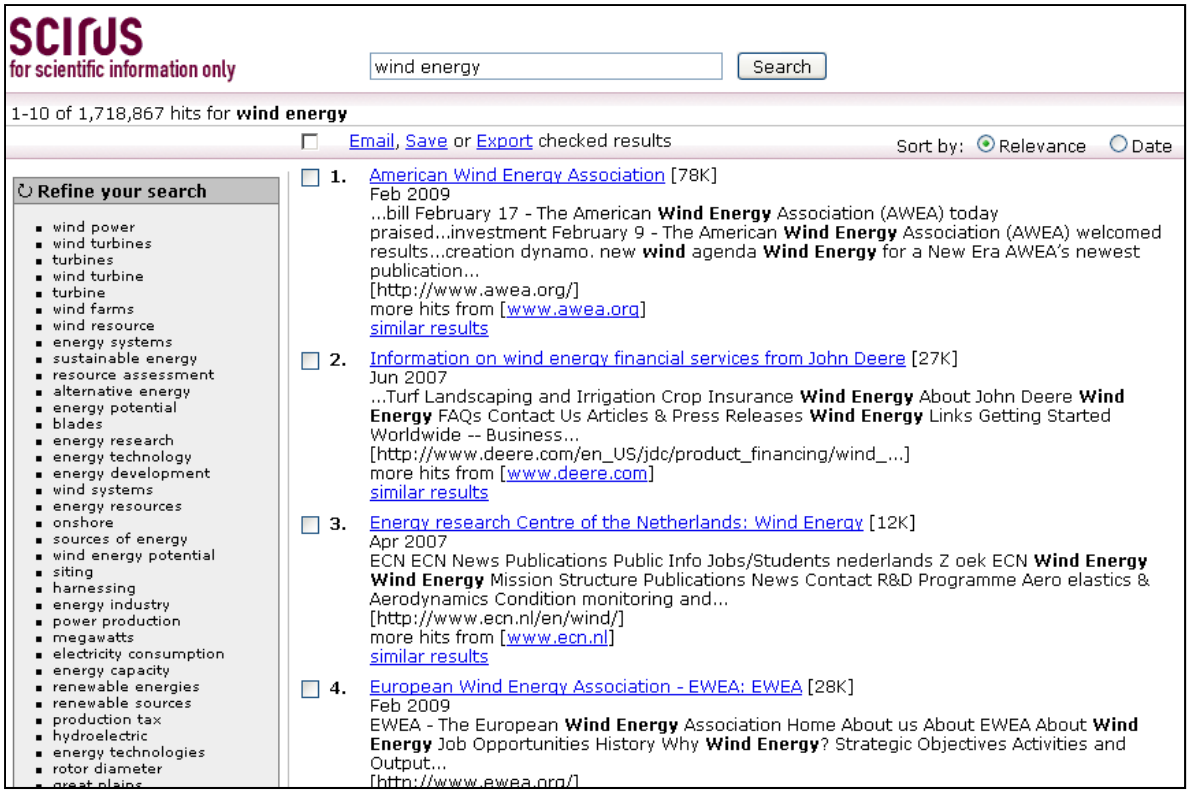

Figure 5. Wind energy was identified as a term related to renewable energy, so in turn it is queried to produce more terms.

Due to the exponential nature of collecting a "tree" of terms in this manner, the number of terms quickly can quickly become unmanageable. For example, at the fourth level in our hierarchy, there are 80^4, or approximately forty-one million terms. In reality, this number tends to be smaller because we find many duplicate terms; as a simple example, *wind energy* includes *wind power* in its list of eighty related terms, and *wind power* also includes *wind energy* in its list.

The breadth (the number of terms on which to perform follow-up searches at each level in the tree) and depth (the number of levels) of the term tree expansion are left as adjustable parameters, to provide control over the terms that are collected. By setting the breadth parameter less than eighty, the term tree can be focused "downward" rather than "outward." This helps to drill down and access highly specific terms, which tend to be more interesting than terms found at high levels of the Scirus hierarchy. For example, terms found at the first level of a search for *renewable energy* include *energy technologies* and *renewable energy sources*. But if we traverse deeper into the tree, we find chains of terms such as *renewable energy -> geothermal -> heat pump -> vapor compression -> accumulator*, at which point we have identified a specific product or technology that is important to a specific subfield of our broader topic. On the other hand, there is a tradeoff involved with focusing on the depth of our term tree rather than the breadth. While it is a positive that terms deep in the tree tend to be narrowly focused and specific, they also have a higher likelihood of being irrelevant to the original topic than terms at higher levels in the tree. For example, another chain of terms we find with renewable energy as the seed is *renewable energy -> polluting -> oligopoly -> marginal cost*, at which point we are entirely outside the realm of our original topic. We attempt to set the breadth and depth parameters appropriately to manage this tradeoff and extract terms that are both specific and relevant. The filtering tools described in sections 3.2.2.2 and 3.2.2.3 of this chapter can also be used to help increase the relevance of the terms collected.

3.2.2.1.1 Software implementation and performance

Scirus term collection is executed by calling the function *get_scirus_terms*, which takes two parameters: a list of seed terms and a list of numbers that represents the breadth and depth pattern of the tree of related terms we wish to collect. The implementation of this function is similar to that of Engineering Village term collection, using Python's URL Library and Regular Expression modules.

For each term in the list of seeds, the URL for the results page is first constructed. This can be done by generating a parameter string which replicates the completion of the Scirus' query form. This allows the relevant search results page to be downloaded, after which regular expressions are used to locate and extract the eighty "refine your search" terms that Scirus provides with the search results. The process is then repeated using some or all of these eighty terms (and in turn, their related terms, etc.), following the pattern represented by the second parameter of *get_scirus_terms*.

This breadth/depth parameter is specified as a list whose length is the number of levels in the tree, and whose elements represent the branching factor at each particular level. So for example, if the list were [20, 10, 5], the first twenty terms would be expanded at the first level of the search, the first ten terms relevant to each term would be expanded in the second level of the search, and the first five

terms would be expanded in the third level of the search. Note that even when a term is not expanded for a follow-up search, it is still added to the list of terms being collected. So in our example, each of the twenty follow-up terms at the first level returns not ten, but eighty terms of its own. Then ten terms related to each of these twenty are further expanded to return eighty of their own. Finally, five terms of each of the ten are expanded to return another eighty each.

Scirus term collection runs in time proportional to the number of search results pages that must be read and parsed. This can be approximated by the product of all the numbers in the extension pattern and the number of seed terms. However, this approximation can be somewhat inaccurate for two reasons. The actual number of pages could be greater because the estimate only accounts for the number of pages opened on the bottom level of the tree. For example, in the pattern [20, 10, 5], with two seed terms, the actual number of pages is $2 + (2 \times 20) + (2 \times 20 \times 10) + (2 \times 20 \times 10 \times 5) = 2242$, whereas our estimate would be $2 \times 20 \times 10 \times 5 = 2000$. On the other hand, the actual number of pages could be less than the estimate due to the fact that we frequently encounter repeated terms, and do not unnecessarily expand them multiple times. For example, a search for *renewable energy* returns *wind energy* and *wind power*. We expand both of these terms, and we find that *wind energy* returns *wind power* as a related term, and similarly, *wind power* returns *wind energy*. Even if the extension pattern calls for these terms to be further expanded, we do not, as this would not produce any new information. This non-duplication of terms tends to significantly lower the number of pages the software must read.

Just as with Engineering Village term collection, the runtime of this module also depends on the connection speed between the local computer and Scirus. As example runtimes under typical conditions, we find that searching with a single seed term and an extension pattern of [10, 5] requires approximately 25 seconds and generates approximately 1,100 terms. Searching with a single seed term with a pattern of [5, 5, 5] requires approximately 40 seconds and generates approximately 1,800 terms.

3.2.2.1.2 Results

Results are presented here for Scirus term extraction, using *renewable energy* as the seed term. We first demonstrate the most basic form, collecting only the eighty terms that Scirus provides when directly queried with *renewable energy*. This is represented by the expansion pattern of [0]. Next, we demonstrate including not only these terms, but also the terms acquired when executing follow-up queries on the first three terms in this list. The expansion pattern for this operation is [3]. Appendix A.2 presents more-substantial term lists generated with the patterns [10] and [2, 2, 2]. In all cases except for the [0] pattern, the list is shorter than the extension pattern would suggest, because duplicate terms are omitted.

Scirus terms, [0] pattern:

- *abatement*
- *alternative energy*
- *alternative fuels*
- *american council*
- *biomass energy*
- *biomass fuels*
- *built environment*
- *climate control*
- *co-operative*
- *commercialisation*
- *distributed generation*
- *efficient appliances*
- *electric energy*
- *electricity generation*
- *electricity markets*
- *electricity sector*
- *electrification*
- *energy crops*
- *energy development*
- *energy industry*
- *energy policy*
- *energy prices*
- *energy research*
- *energy resources*
- *energy services*
- *energy systems*
- *energy technologies*
- *energy technology*
- *environmental costs*
- *ethanol*
- *frame-based*
- *fuel cell*
- *fuel cells*
- *fuel economy*
- *future energy*
- *geothermal*
- *geothermal energy*
- *geothermal power*
- *greenhouse gas abatement*
- *greenhouse gas emissions*
- *harnessing*
- *hydro power*
- *hydroelectric*
- *hydrogen fuel*
- *impact on the environment*
- *intelligent software*
- *landfill gas*
- *national security*
- *non-renewable*
- *photovoltaic*
- *photovoltaics*
- *polluting*
- *rated power*
- *renewable energies*
- *renewable energy resources*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable resource*
- *renewable sources*
- *rural electrification*
- *rural energy*
- *solar photovoltaics*
- *solar power*
- *sources of energy*
- *stand-alone*
- *state energy*
- *sustainable energy*
- *tax incentives*
- *tidal power*
- *turbine*
- *turbines*
- *wave energy*
- *wind energy*
- *wind farms*
- *wind generation*
- *wind power*
- *wind resource*
- *wind turbine*
- *wind turbines*

Scirus terms, [3] pattern:

- *abatement*
- *alternative energy*
- *alternative fuels*
- *american council*
- *ashrae*
- *atmospheric sciences*
- *battery consortium*
- *biodiesel*
- *bioenergy*
- *biogas*
- *biological attack*
- *biomass*
- *biomass combustion*
- *biomass energy*
- *biomass fuels*
- *biomass gasification*
- *building energy*
- *built environment*
- *capital cost*
- *carbon emissions*
- *clean coal*
- *climate control*
- *climate policy*
- *co-operative*
- *commercialisation*
- *conversion systems*
- *cost reduction*
- *data center*
- *decision-support system*
- *delphi study*
- *development mechanism*
- *diesel*
- *digester gas*
- *distributed generation*
- *distribution of electricity*
- *domestic energy*
- *domestic production*
- *dual fuel engine*
- *economic costs*
- *efficient appliances*
- *electric energy*
- *electric grid*
- *electric power systems*
- *electricity consumption*
- *electricity demand*
- *electricity generation*
- *electricity markets*
- *electricity production*
- *electricity sector*
- *electrification*
- *electroplating*
- *energy crops*
- *energy development*
- *energy economics*
- *energy efficiency*
- *energy farming*
- *energy impacts*

- *energy industry*
- *energy management*
- *energy policy*
- *energy potential*
- *energy prices*
- *energy products*
- *energy research*
- *energy resource*
- *energy resources*
- *energy savings*
- *energy sector*
- *energy services*
- *energy supplies*
- *energy systems*
- *energy technologies*
- *energy technology*
- *energy utilization*
- *energy-efficient lighting*
- *environmental costs*
- *ethanol*
- *experience curves*
- *fall in price*
- *form of energy*
- *frame-based*
- *fuel cell*
- *fuel cells*
- *fuel economy*
- *future energy*
- *gasoline*
- *geothermal*
- *geothermal energy*
- *geothermal power*
- *greenhouse gas abatement*
- *greenhouse gas emissions*
- *harmful emissions*
- *harnessing*
- *heat pumps*
- *hydro power*
- *hydroelectric*
- *hydrogen fuel*
- *hydropower*
- *impact on the environment*
- *industrial technologies*
- *installed capacity*
- *insulation*
- *integrated assessment*
- *intelligent software*
- *intermittent*
- *investor*
- *irrigation water pumping*
- *landfill gas*
- *lead-acid batteries*
- *market barriers*
- *metal recovery*
- *milieukunde*
- *mitigation*
- *national security*
- *non-renewable*
- *office of basic energy sciences*
- *office of science*
- *option to purchase*
- *photovoltaic*
- *photovoltaic cells*
- *photovoltaic systems*
- *photovoltaics*
- *polluting*
- *poultry litter*
- *power electronics*
- *producer gas*
- *propulsion*
- *public affairs*
- *rated power*
- *reducing greenhouse gas emissions*
- *renewable energies*
- *renewable energy*
- *renewable energy resources*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable resource*
- *renewable sources*
- *republic of croatia*
- *resource assessment*
- *ress*
- *rinse water*
- *role of particles*
- *rural electrification*
- *rural electrification in senegal*
- *rural energy*
- *rural livelihoods*
- *scheikunde*
- *senegal*
- *small island developing states*
- *solar collectors*
- *solar electricity*
- *solar energy conversion*
- *solar energy utilization*
- *solar photovoltaics*
- *solar power*
- *solar water*
- *sources of energy*
- *stand-alone*
- *state energy*
- *storage technologies*
- *storage technology*
- *sustainable development*
- *sustainable energy*
- *sustainable future*
- *tax incentives*
- *technological learning*
- *technology choice*
- *technology transfer*
- *thermal power*
- *thermal systems*
- *tidal power*
- *transportation fuels*
- *turbine*
- *turbines*
- *united nations development program*
- *waste heat*
- *water heating*
- *water pumping*
- *water turbine*
- *wave energy*
- *welding*
- *wind energy*
- *wind farms*
- *wind generation*
- *wind power*
- *wind resource*
- *wind turbine*
- *wind turbines*
- *world bank*

The terms extracted using the [0] pattern (the eighty "refine your search" terms of *renewable energy* itself) contain a large proportion of highly relevant terms. With the exception of a small number of terms such as *stand-alone* and *co-operative*, most terms rightfully belong. It can also be seen that most of renewable energy's broad subfields are represented on this list. We see, for example, *biomass energy*, *geothermal energy*, *fuel cells*, *geothermal energy*, *hydro power*, *hydrogen fuel*, *solar power*, *wave energy*, and *wind energy*. On the other hand, the terms in this list tend to be overly general, which we would expect since they are found only one level down in the hierarchy from *renewable energy* itself.

With the [3] pattern, three terms are expanded into the next level of the hierarchy, attempting to increase the number of terms that represent narrow pieces of the renewable energy field. We find that the list does indeed contain terms such as *biodiesel*, *biomass gasification*, *clean coal*, *electroplating*, *energy-efficient lighting*, and *rural electrification in senegal*, demonstrating more specificity than the original list. However, traversing the hierarchy in this way can lead to a larger proportion of irrelevant terms. This list, for example, contains the terms *investor*, *office of science*, *option to purchase*, and *small island developing states*. We must manage the tradeoff between relevance and specificity, but as with the choice between Engineering Village controlled or uncontrolled terms, we have a preference for traversing deeper in the hierarchy and accepting the risk of irrelevant terms.

3.2.2.2 Relevance testing by back-pointing

In our basic approach to Scirus keyword extraction, we view Scirus' "refine your search" terms as representing a hierarchy, or tree, of keywords. The hierarchy is rooted at our seed term, and directly beneath each term in the hierarchy are the eighty terms Scirus suggests when queried with that term. This tree is explored as widely and deeply as desired, and all non-duplicate terms encountered on this traversal are extracted.

This approach works reasonably well, but we find that in practice, even with the ability to vary the breadth/depth parameters, the extracted keywords tend to include too many unsuitable terms, which tend to fall into two main categories: They can either be overly general or obvious, as in *energy technologies* or *renewable energy sources*, or they can be irrelevant to the original seed term, as in *marginal cost*. Unfortunately, solutions to these two problems tend to be conflicting. The first problem, over-generality, can be addressed by increasing the depth of our tree while decreasing its breadth, while the second problem, irrelevance, is addressed in the opposite way.

Relevance filtering by back-pointing is an attempt at circumventing this apparent opposition between specificity and relevance by first collecting terms that are unlikely to be overly general, i.e. collecting them from deep in the term hierarchy, and next, identifying and removing terms from this list that are irrelevant to the original topic. This section describes our back-pointing method of identifying the irrelevant terms for removal.

The concept of back-pointing is straightforward yet surprisingly effective. We say that term A "points" to term B if B is one of the eighty "refine your search" terms Scirus provides when queried with term A. Under this terminology, the basic keyword extraction methodology is to collect all the terms to

which the seed term points, all the terms to which those term point, etc. Back-pointing relevance checking imposes a restriction on the terms we extract: they must point *back* to the original seed term. For example, using *renewable energy* as our seed term, one of the terms it points to is *turbines*. When *turbines* is then queried, we find that two of the terms it points to are *microhydro* and *horizontal axis*. Of these two terms, we find that *microhydro* points back to *renewable energy*, while *horizontal axis* does not. This would lead us to conclude that *microhydro* is relevant to *renewable energy* even though it is found two levels down in the term hierarchy, whereas *horizontal axis*, while relevant to *turbines*, is not directly relevant to the larger field of renewable energy.

It can be overly restrictive to require all terms to point directly back to the seed term. In the example above, *horizontal axis* was filtered out because it does not point back to *renewable energy*. However, we find that it does point back to *turbines*, and *turbines* in turn points back to *renewable energy*. Because *horizontal axis* points to a term which points to the seed term, we say that it points back through two levels. It is often the case that we wish to keep such a term. (Note that it is not always the case, as it was in this example, for term B to point to term A whenever term A points to term B. Here we had that *renewable energy* points to *turbines*, and vice versa, and we also had that *turbines* points to *horizontal axis*, and vice versa. But as a counterexample, *renewable energy* also points to *geothermal*, which does not point back to *renewable energy*, and similarly, *turbines* also points to *optimization*, which does not point back to turbines.)

The maximum number of levels through which a term may point back to the seed term in order to be considered relevant is an adjustable parameter. In the limiting case, if we kept terms that point back through any number of levels, this would amount to reverting to our basic Scirus keyword extraction, with no filtering. Two levels tends to be a good choice, striking a balance between removing completely irrelevant terms and keeping marginally relevant ones.

The diagram below demonstrates the use of back-pointing.

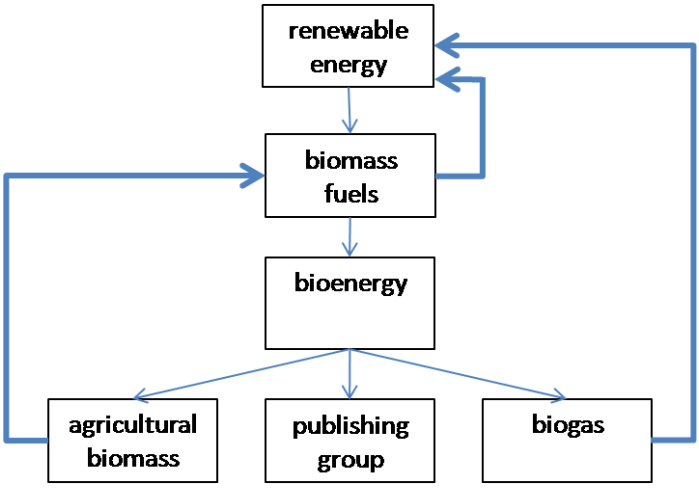

Figure 6. All Scirus connections among the terms *renewable energy*, *biomass fuels*, *bioenergy*, *agricultural biomass*, *publishing group*, and *biogas* are represented. The thicker lines represent back-pointing paths to renweable energy. *Biomass fuels* and *biogas* point directly back to *renewable energy*. *Bioenergy* and *agricultural biomass* point back through two levels. The irrelevant term *publishing group* has no back-pointing connection to *renewable energy*.

There are two variants to our back-pointing methods. In one version, the user provides a list of keywords which have been produced using the previously described term collection algorithms, and the software checks each of these keywords to see if they point back to the seed term. Only those that do point back are kept. We call this method "back-pointing term filtering." This method works well for simple applications, but has two drawbacks. First, it is often the case that the resultant filtered list is quite short; especially in the case of *renewable energy*, we find that a very small proportion of terms to which *renewable energy* points point back to *renewable energy*. Second, this method requires significant runtime. Scirus must be queried on every term in the list to identify their eighty related keywords, and each of these related lists must be scanned to see if they contain the seed term.

Due to these drawbacks with directly looking up whether each term points back to the seed term, we have developed a second back-pointing method, which we call "back-pointing term collection." In this method, instead of being filtered from an input list, back-pointing terms are produced with no input. The process involves having first constructed a local, offline database that represents a subset of the Scirus term tree. Then, at any point in the future, this database can be queried to produce all terms pointing back to the seed, through however many levels we choose to allow. The preliminary step of this process, creating a local database, requires an extremely large runtime, but this only needs to be done once before the database can be utilized an unlimited number of times. The second step of the process is very fast, since it only involves querying a local database. Given a large enough database, on the order of thousands of terms, we can nearly instantly produce a large number of back-pointing terms.

3.2.2.2.1 Software implementation and performance

The implementation of the filtering variant of back-pointing testing utilizes much of the functionality from Scirus term collection. The function *filter_for_back_pointing_terms* takes two parameters: the list of terms to be filtered and the seed term. For each term in the list, it uses the Scirus term collection code to identify its eighty related keywords. If any of these related keywords are the seed term, the software identifies the term being queried as a back-pointing term.

This function runs in time proportional to the number of terms in the list being filtered. The performance bottleneck is the time required to open and read a search results page for each term. Under typical network conditions, filtering 100 terms requires approximately 75 seconds, and filtering 200 terms requires approximately 150 seconds.

The implementation of the term generation variant of back-pointing testing requires first creating a local database that represents a portion of the term connections inherent in Scirus. Python's SQLite DB-API2 interface is used to create the database. This SQL database consists of a single table with two columns. The first column is the name of a term. The second column is a comma-separated list of terms that point to the term in the first column. This structure is an incoming adjacency list representation of the Scirus term hierarchy. Our software populates this database by exploring Scirus' terms in a breadth-first manner, beginning with the seed term.

Beginning with the seed term, our Scirus term collection code is used to identify its eighty directly related terms. We add a record to our database for each of these eighty terms, where the first field consists of the term and the second field consists of the seed term (the term which pointed to it). We then proceed to expand each of these terms to identify their eighty directly related terms, and continue this breadth-first exploration for as long as desired. For each term encountered, if it had not previously been encountered, a new record is added to the database, using that term as the first field, and the term that pointed to it as the second field. If, on the other hand, the term had been encountered before, we locate its previously existing record in the database, and augment the second field by appending the term that was most recently discovered to point to it.

The process of exploring Scirus' terms to create a sizeable local database is time-consuming. Furthermore, this database can be made arbitrarily large, and can therefore take an arbitrarily long time to construct. For these reasons, we make it possible for our software to stop running and then continue where it left off at a later time. In this way, we can have it continually expand our database by a manageable amount each time we decide to run it. The breadth-first search of Scirus terms requires a so-called "visited list" and an "unvisited list." The visited list contains terms which have been expanded, i.e. terms for which we have identified their eighty related terms. The unvisited list contains terms that we have discovered but have not yet expanded. The process consists of selecting the term at the front of the unvisited list, moving it to the visited list, identifying its eighty related terms, and putting these at the end of the unvisited list. In order to allow our software to continue its search where it left off, we write the visited and unvisited lists to disk at the end of a session, and read them at the beginning of the next.

Producing a local database of terms is one part of the process of producing back-pointing terms. The other part consists of querying this database to identify all of the terms which point back to the seed. Due to the structure of our database, this is quite simple and efficient. Our software simply queries the database for the second field of the record for which the first field is the seed term. In other words, we find the record which represents our seed term, and extract from it the comma-separated list of terms that point to that seed term. An example SQL query is *select in_pointing_terms from scirus_terms where term='renewable energy'*.

A direct query of this particular row of our database gives the terms that point back to the seed through one level. If we wish to identify terms which point back through more than one level, we then perform follow-up queries on all of the terms which were identified as pointing back through one level. For example, if we are looking for terms that point back through two levels, we first query the database to find terms that point back through one level, and then query the database again for terms that point back to each of these.

The software that creates the database of term relations runs in time roughly proportional to the square of the number of terms we wish to expand. Initially, the bottleneck involves reading a term's search results page and collecting its eighty related terms. However, another component of the runtime is the process of adding the eighty related terms to our SQL database. For each of these eighty terms, the entire database must be searched to determine whether a record for the term already exists. When

the database is small, the time required for this operation is negligible. However, as the database grows, this becomes more and more time-intensive. The runtime is therefore $O(n^2)$, since for each term, all previously encountered terms must be searched. However, until the local database become exceedingly large, querying Scirus for a term's search results page is a much slower operation than querying the local database for instances of that term, so for a small number of terms, the runtime is roughly linear in n.

As example runtimes, extending 50 terms given an initially empty database requires approximately 90 seconds. Extending 50 terms given a database with 60,000 terms, however, requires approximately 10 minutes. Extending 100 terms given an initially empty database requires approximately 175 seconds, and extending 100 terms given a database with 60,000 terms requires approximately 20 minutes.

3.2.2.2. Results

Results are presented here for relevance checking by back-pointing. First we demonstrate the filtering variant of our back-pointing methodology. We use the terms extracted from Scirus using the [3] pattern, as demonstrated in section 3.2.2.1.2. We display the subset of these terms that point back to *renewable energy*, as well as the subset that do not. Next, we demonstrate the term generation variant. Here, we use the first 250 terms acquired in a breadth-first search of Scirus' term hierarchy from the term *renewable energy* as our potential terms. We display all such terms that point back to *renewable energy* through one, two, or three levels. In Appendix A.3, we present term generation back-pointing results with a larger set of potential terms; namely, the first 1,000 terms acquired in a breadth-first search of Scirus' term hierarchy.

Term filtering by back-pointing:

Back-pointing terms:

- *biogas*
- *biomass fuels*
- *conversion systems*
- *electricity production*
- *electrification*
- *energy crops*
- *energy supplies*
- *experience curves*
- *hydropower*
- *irrigation water pumping*
- *market barriers*
- *photovoltaics*
- *renewable energies*
- *renewable sources*
- *rural electrification*
- *rural energy*
- *tidal power*
- *turbine*
- *turbines*
- *wave energy*
- *wind farms*
- *wind turbine*
- *wind turbines*

Non-back-pointing terms:

- *abatement*
- *alternative energy*
- *alternative fuels*
- *american council*
- *ashrae*
- *atmospheric sciences*
- *battery consortium*
- *biodiesel*
- *bioenergy*
- *biological attack*
- *biomass*
- *biomass combustion*
- *biomass energy*
- *biomass gasification*
- *building energy*
- *built environment*
- *capital cost*
- *carbon emissions*

- *clean coal*
- *climate control*
- *climate policy*
- *co-operative*
- *commercialisation*
- *cost reduction*
- *data center*
- *decision-support system*
- *delphi study*
- *development mechanism*
- *diesel*
- *digester gas*
- *distributed generation*
- *distribution of electricity*
- *domestic energy*
- *domestic production*
- *dual fuel engine*
- *economic costs*
- *efficient appliances*
- *electric energy*
- *electric grid*
- *electric power systems*
- *electricity consumption*
- *electricity demand*
- *electricity generation*
- *electricity markets*
- *electricity sector*
- *electroplating*
- *energy development*
- *energy economics*
- *energy efficiency*
- *energy farming*
- *energy impacts*
- *energy industry*
- *energy management*
- *energy policy*
- *energy potential*
- *energy prices*
- *energy products*
- *energy research*
- *energy resource*
- *energy resources*
- *energy savings*
- *energy sector*
- *energy services*
- *energy systems*
- *energy technologies*
- *energy technology*
- *energy utilization*
- *energy-efficient lighting*
- *environmental costs*
- *ethanol*
- *fall in price*
- *form of energy*
- *frame-based*
- *fuel cell*
- *fuel cells*
- *fuel economy*
- *future energy*
- *gasoline*
- *geothermal*
- *geothermal energy*
- *geothermal power*
- *greenhouse gas abatement*
- *greenhouse gas emissions*
- *harmful emissions*
- *harnessing*
- *heat pumps*
- *hydro power*
- *hydroelectric*
- *hydrogen fuel*
- *impact on the environment*
- *industrial technologies*
- *installed capacity*
- *insulation*
- *integrated assessment*
- *intelligent software*
- *intermittent*
- *investor*
- *landfill gas*
- *lead-acid batteries*
- *metal recovery*
- *milieukunde*
- *mitigation*
- *national security*
- *non-renewable*
- *office of basic energy sciences*
- *office of science*
- *option to purchase*
- *photovoltaic*
- *photovoltaic cells*
- *photovoltaic systems*
- *polluting*
- *poultry litter*
- *power electronics*
- *producer gas*
- *propulsion*
- *public affairs*
- *rated power*
- *reducing greenhouse gas emissions*
- *renewable energy*
- *renewable energy resources*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable resource*
- *republic of croatia*
- *resource assessment*
- *ress*
- *rinse water*
- *role of particles*
- *rural electrification in senegal*
- *rural livelihoods*
- *scheikunde*
- *senegal*
- *small island developing states*
- *solar collectors*
- *solar electricity*
- *solar energy conversion*
- *solar energy utilization*
- *solar photovoltaics*
- *solar power*
- *solar water*
- *sources of energy*
- *stand-alone*
- *state energy*
- *storage technologies*
- *storage technology*
- *sustainable development*
- *sustainable energy*
- *sustainable future*
- *tax incentives*
- *technological learning*
- *technology choice*
- *technology transfer*
- *thermal power*
- *thermal systems*
- *transportation fuels*
- *united nations development program*
- *waste heat*
- *water heating*
- *water pumping*
- *water turbine*
- *welding*
- *wind energy*
- *wind generation*
- *wind power*
- *wind resource*
- *world bank*

The results show that back-pointing for term filtering does indeed increase the proportion of relevant terms. The filtered list contains very few irrelevant terms. *Experience curves* does not belong, and *market barriers* is only somewhat relevant, but the rest of the terms on the list are pertinent to renewable energy. In contrast, the list of terms that were removed contains a large proportion of irrelevant terms; for example, we see *co-operative*, *commercialization*, *data center*, *development mechanism*, *intermittent*, and *rinse water*.

However, the list of filtered terms also contains many valuable terms that should not have been labeled irrelevant. A small subset of the highly relevant terms that were filtered out includes *alternative energy*, *alternative fuels*, *biomass*, *fuel cell*, *geothermal*, and *wind power*. These results show that single-level term filtering may be too restrictive. The next set of results demonstrates term generation, as opposed to filtering, using back-pointing through one, two, or three levels.

Term generation by back-pointing, 250 potential terms, levels 1, 2, and 3:

Level 1

- *biomass fuels*
- *wind power development*
- *turbines*
- *experience curves*
- *electrification*
- *turbine*
- *electricity production*
- *kilowatts*
- *renewable sources*
- *megawatts*
- *hydroelectric*
- *water heating*
- *biogas*
- *wave energy*
- *wind energy potential*
- *energy supplies*
- *renewable energies*
- *rural energy*
- *energy impacts*
- *wind turbines*
- *tidal power*
- *energy crops*
- *rural electrification*
- *power development*
- *wind farms*
- *wind*
- *wind turbine*

Level 2

- *alternative energy*
- *alternative fuels*
- *bioenergy*
- *biofuel*
- *biogas plants*
- *biomass*
- *biomass combustion*
- *biomass energy*
- *biomass gasification*
- *blades*
- *carbon emissions*
- *clean coal*
- *conversion systems*
- *development mechanism*
- *diesel*
- *digester gas*
- *distributed generation*
- *distribution of electricity*
- *domestic energy*
- *domestic production*
- *efficient appliances*
- *electric energy*
- *electric power systems*
- *electricity consumption*
- *electricity demand*
- *electricity generation*
- *electricity markets*
- *electricity sector*
- *electricity supply*
- *energy capacity*
- *energy development*
- *energy efficiency*
- *energy industry*
- *energy output*
- *energy policy*
- *energy potential*
- *energy prices*
- *energy products*
- *energy research*
- *energy resource*
- *energy resources*
- *energy savings*
- *energy sector*
- *energy services*
- *energy systems*
- *energy technologies*
- *energy technology*
- *energy utilization*
- *environmental benefits*
- *farm operation*
- *fuel cell*
- *future energy*
- *gasification*
- *geothermal*
- *geothermal energy*
- *geothermal power*
- *greenhouse gas emissions*
- *harnessing*
- *heat pumps*
- *hydro power*

- *hydroelectric dams*
- *hydrogen fuel*
- *installed capacity*
- *irrigation water pumping*
- *landfill gas*
- *market barriers*
- *municipal utilities*
- *non-polluting*
- *non-renewable*
- *onshore*
- *photovoltaic*
- *photovoltaic systems*
- *photovoltaics*
- *power production*
- *power sources*
- *production tax*
- *propulsion*
- *rated power*
- *reducing greenhouse gas emissions*
- *renewable energy resources*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable resource*
- *roaring forties*
- *rotor blades*
- *rotor diameter*
- *scheikunde*
- *siting*
- *solar cooker*
- *solar electricity*
- *solar energy utilization*
- *solar photovoltaics*
- *solar power*
- *solar water*
- *sources of energy*
- *state energy*
- *storage technologies*
- *sustainable energy*
- *technological learning*
- *thermal power*
- *thermal systems*
- *tidal energy*
- *turbine blades*
- *waste heat*
- *water pumping*
- *water turbine*
- *wind energy*
- *wind events*
- *wind generation*
- *wind power*
- *wind power plant*
- *wind resource*
- *wind speed data*
- *wind system*
- *wind system*

Level 3

- *american council*
- *ashrae*
- *battery consortium*
- *biodiesel*
- *building energy*
- *built environment*
- *capital cost*
- *carbon sequestration*
- *climate control*
- *cost reduction*
- *doe program*
- *dual fuel engine*
- *energy management*
- *energy-efficient lighting*
- *environmental costs*
- *environmental education*
- *ethanol*
- *financial incentives*
- *form of energy*
- *fuel cells*
- *fuel charge*
- *fuel economy*
- *gasoline*
- *greenhouse gas abatement*
- *harmful emissions*
- *impact on the environment*
- *industrial technologies*
- *insulation*
- *integrated assessment*
- *integrated resource*
- *lead-acid batteries*
- *learning rates*
- *local residents*
- *long-term contracts*
- *maldives*
- *market simulation*
- *metal recovery*
- *miles per hour*
- *mitigation*
- *national security*
- *office of basic energy sciences*
- *office of science*
- *polluting*
- *power electronics*
- *producer gas*
- *public affairs*
- *republic of croatia*
- *resource assessment*
- *ress*
- *rice hulls*
- *role of particles*
- *small island developing states*
- *solar energy conversion*
- *storage technology*
- *sustainable future*
- *system voltage*
- *tariff*
- *tax credit*
- *tax incentives*
- *transportation fuels*
- *utilization factor*
- *vice president*
- *warm springs*
- *wind velocity*

Of the terms we have generated that point back to *renewable energy* through one level, many are the same as those displayed in the filtering results. They are highly relevant, but the list lacks many other equally relevant terms. The addition of terms that point back through two levels remedies the situation. It was noted in the filtering results that the following terms were not included: *alternative energy*, *alternative fuels*, *biomass*, *fuel cell*, *geothermal*, and *wind power*. Here we see that all of them have been identified as two-level back-pointing terms. Furthermore, the results show that the introduction of two-level back-pointing terms does not significantly increase the proportion of irrelevant terms. Of the two-level terms listed here, very few are irrelevant. *Roaring forties* and *siting* perhaps should not be on the list, but for the most part, the terms are quite pertinent.

When three-level back-pointing terms are included, the terms begin to show a lower degree of relevance. The list does contain some valuable terms such as *biodiesel*, *fuel cells*, and, *solar energy conversion*, but it also contains irrelevant ones such as *learning rates*, *miles per hour*, *national security*, *market simulation*, and *vice president*. It should be noted that of the 250 candidate terms from our Scirus breadth-first search, 207 point back to *renewable energy* through one, two, or three levels. Including terms from all three sets therefore produces similar results to simply using all terms encountered in the breadth-first search. Due to the irrelevant terms produced by three-level back pointing, we typically only use terms that point back through one or two levels.

3.2.2.3 Relevance testing by eigenvector centrality

The back-pointing approach to relevance filtering described in the previous section is relatively straightforward. Only terms that point back to the seed term within a small number of levels are retained. This is a good heuristic, but relevance filtering can be made more robust with a more complex approach. For example, if a term just exceeds the threshold for number of levels through which it points back to the seed, but is densely connected with many other terms that have already been declared relevant, it would make sense to declare that term relevant as well. In this section, we describe an approach to relevance filtering that considers a term's connections with all other terms, instead of only considering its connection to the seed term. It is a mathematically rigorous approach that makes use of the so-called eigenvector centrality metric.

So far we have discussed a tree-like structure of terms identified by Scirus. A search for a particular term in Scirus results in eighty new terms being identified, so we can consider each term as a node in a tree with eighty outgoing directed edges to other terms. This tree-like, hierarchical view is useful for the back-pointing relevance checking described in the previous section. But a more accurate view is not of a *tree*, but of a densely connected *network* of terms, with the difference being that a tree is by definition acyclic and has exactly one parent node for each non-root node. As an example of the structure of a Scirus term network, consider the small set of terms *renewable energy*, *renewable energy technologies*, *wind power*, and *geothermal*. A search for *renewable energy* returns each of the other three as related terms, so we represent this in our network with edges pointing from renewable energy to each of the other three terms. A search for *renewable energy technologies* returns both *wind power* and *geothermal*, so we add such edges to our network. A *wind power* search returns *renewable energy technologies* and *geothermal*, so again we add the appropriate edges. Finally, a search for *geothermal*

does not return any of the other three terms, so *geothermal* has no outgoing edges in our network. This network contains cycles (e.g. *renewable energy technologies* -> *wind power* -> *renewable energy technologies*) and has nodes with more than one parent node (e.g. *geothermal* is linked to by all three other terms), so it has a network or web-like structure rather than a tree structure. This is illustrated in the diagram below.

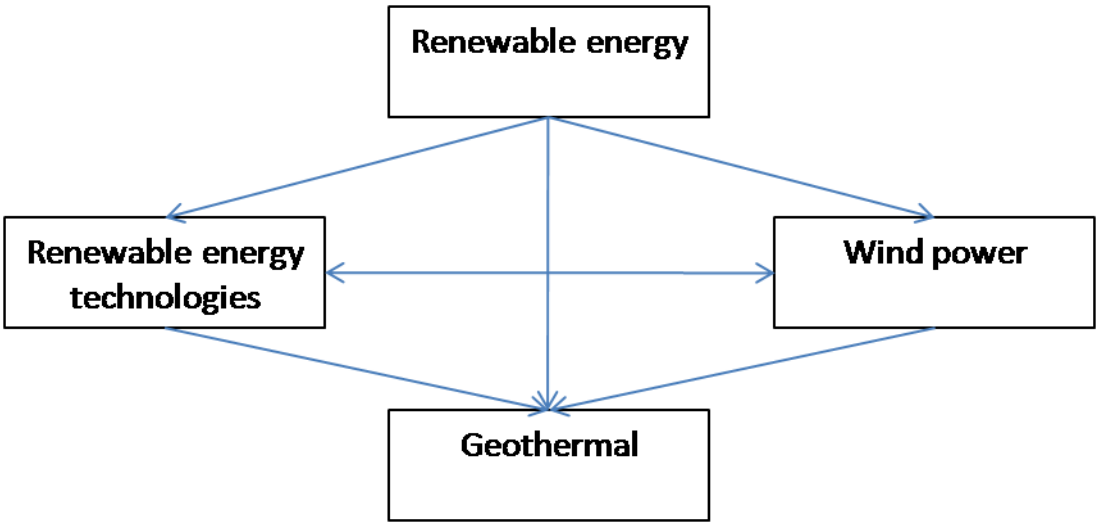

Figure 7. Term network representing the relationships among the terms *renewable energy*, *renewable energy technologies*, *wind power*, and *geothermal*.

Given a network of terms, we can then calculate the eigenvector centrality of each term, or node, in our network. It is a measure of node importance, where the centrality of a node X is defined as the sum of the centralities of the nodes with edges pointing to X, divided by a constant. Intuitively, if X has many incoming edges, and particularly if those edges come from nodes with large centralities, then X will have a large centrality. Eigenvector centrality is illustrated conceptually in the figure below.

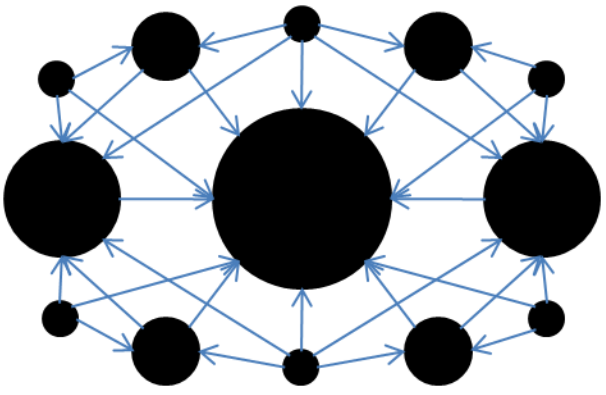

Figure 8. Eigenvector centrality depiction of a generic network. The centrality of a node is defined in terms of the number of incoming edges to that node, as well as the centralities of the nodes from which these edges originate. Here, the displayed size of a node represents its centrality.

In order to use the eigenvector centrality metric for relevance checking, we must first create a representation of the network in which we are interested. Scirus has a vast underlying term network,

only a very small subset of which is related to our topic. So we create a subnetwork of interest by beginning with a seed term, e.g. *renewable energy*, and exploring Scirus' term network in a breadth-first manner, storing the nodes and edges we discover in a local database to be used for eigenvector centrality analysis. The stored subnetwork is continually expanded until it is sizeable enough that it is unlikely to be missing many important terms. At that point, we perform the calculations on this network and rank each term according to its eigenvector centrality. Lower-ranked terms are then declared irrelevant.

It is important to note that this eigenvector centrality calculation does not explicitly test a given term for relevance with the seed term. It merely tests the term's centrality, i.e. importance, of the term in a network *centered on* the seed term. The intuition behind the use of this metric is that terms that are relevant to the seed term will be closely tied to other terms in this network, and conversely, terms that are irrelevant to the seed term will be only sparsely connected with the greater network.

3.2.2.3.1 Software implementation and performance

The eigenvector centrality implementation requires a local database that represents Scirus' underlying network of terms. This process is nearly identical to that of producing a database for use in back-pointing term generation, with one important difference. For back-pointing, we are interested in all terms encountered in a breadth-first search of the Scirus term network, including those terms which have not yet been extended, i.e. those which are on the unvisited list. For eigenvector centrality, on the other hand, we are only interested in terms for which outgoing links have been identified, i.e. terms that are on the visited list. The reason for this is that we do not want to evaluate the centrality of terms which are artificially represented as "sink" nodes in our network.

Because of this difference, the database is structured as an outgoing adjacency list, rather than an incoming adjacency list. The first field in the table contains the term itself, while the second field stores a comma-separated list of the terms pointed to by that term. When conducting a breadth-first search of the Scirus term network to construct our database, a record is added for each term processed, with its eighty outgoing neighbors listed in the second field of the table. This means that every term with a record in the table is part of the visited list; terms on the unvisited list are contained in the table as second-field values, but will not appear in any records as first-field values.

Given our database representation of a subset of the Scirus term network, we begin the eigenvector centrality calculations by creating a term adjacency matrix in Python. By definition, an adjacency matrix A contains values $A_{i,j}$ such that $A_{i,j} = 1$ if term i points to term j, and $A_{i,j} = 0$ otherwise. Our software creates this matrix by directly reading the rows of our database.

After creating the term adjacency matrix, our software uses Python's linear algebra package to calculate the eigenvalues and eigenvectors of this matrix. Only the principal eigenvector, i.e. that which has the largest associated eigenvalue, is needed. Each value V_i in this principal eigenvector represents the centrality of term i.

The runtime of the database creation is proportional to the number of terms extracted. In contrast to database creation for back-pointing, each term extracted can be immediately added to the database without needing to check if the term has already been encountered. Starting from an empty database, compiling a 50 term collection takes approximately 40 seconds while collecting 50 terms given a database with 3,000 previous terms requires approximately 45 seconds. Extending 100 terms given an initially empty database requires approximately 85 seconds, and extending 100 terms given a database with 3,000 extended requires approximately 90 seconds.

Given a local database, the runtime of the eigenvector centrality calculation is approximately proportional to n^3, the cube of the number of terms we are analyzing. Creating the adjacency matrix requires $O(n^2)$ time, but since it only involves querying a local database, this operation is nearly instantaneous. Eigenvector calculation is the most significant bottleneck. Calculating the eigenvectors of an $n \times n$ matrix requires $O(n^3)$ time in the best case. But if the matrix becomes too large, the software is unable to perform the calculation entirely in RAM, and must perform memory swaps between the computer's RAM and the hard drive. This significantly reduces the speed of the calculation.

Calculating term centralities on a network of 250 terms takes approximately 1 second. For a network of 1,000 terms, it takes approximately 55 seconds. But due to the large asymptotic growth of the runtime of eigenvector calculation, a network of 5,000 terms requires approximately 4 hours.

3.2.2.3.2 Results

In this section, results are presented for eigenvector centrality relevance testing. Using a network consisting of the first 250 terms encountered in a breadth-first search of Scirus from the term *renewable energy*, a ranked list of all terms by eigenvector centrality is displayed. In Appendix A.5, results are displayed for a 1,000-term network.

Eigenvector centrality rankings, 250-term network:

- *renewable energy*
- *renewable energy sources*
- *renewable energy technologies*
- *energy technologies*
- *wind energy*
- *renewable sources*
- *rural energy*
- *renewable energy resources*
- *renewable energy systems*
- *renewable energies*
- *wind power*
- *energy resource*
- *energy development*
- *alternative energy*
- *sources of energy*
- *energy resources*
- *wind power development*
- *wind energy potential*
- *energy systems*
- *energy technology*
- *solar electricity*
- *sustainable energy*
- *wind resource*
- *rural electrification*
- *biomass energy*
- *hydro power*
- *future energy*
- *energy policy*
- *state energy*
- *wind*
- *electric energy*
- *energy industry*
- *energy impacts*
- *wind systems*
- *energy sector*
- *geothermal energy*
- *energy supplies*
- *wind turbines*
- *electricity production*
- *distributed generation*
- *solar photovoltaics*
- *energy research*
- *electricity supply*
- *energy efficiency*
- *solar power*
- *energy services*
- *wind farms*
- *photovoltaic systems*
- *energy capacity*
- *wave energy*

- *megawatts*
- *energy potential*
- *power development*
- *wind turbine*
- *wind generation*
- *electricity sector*
- *solar energy utilization*
- *tidal power*
- *geothermal power*
- *experience curves*
- *energy products*
- *turbine*
- *installed capacity*
- *electricity generation*
- *tidal energy*
- *electrification*
- *domestic energy*
- *energy management*
- *bioenergy*
- *turbines*
- *hydrogen fuel*
- *building energy*
- *clean coal*
- *production tax*
- *kilowatts*
- *fuel cell*
- *power sources*
- *wind system*
- *development mechanism*
- *photovoltaics*
- *rated power*
- *solar energy conversion*
- *renewable resource*
- *hydroelectric*
- *form of energy*
- *technological learning*
- *market barriers*
- *storage technologies*
- *distribution of electricity*
- *conversion systems*
- *energy-efficient lighting*
- *geothermal*
- *energy output*
- *solar water*
- *harnessing*
- *energy utilization*
- *energy savings*
- *biomass*
- *energy crops*
- *fuel cells*
- *municipal utilities*
- *efficient appliances*
- *electricity demand*
- *electric power systems*
- *wind speed data*
- *wind power plant*
- *energy prices*
- *alternative fuels*
- *water pumping*
- *landfill gas*
- *non-renewable*
- *rotor diameter*
- *irrigation water pumping*
- *reducing greenhouse gas emissions*
- *resource assessment*
- *water heating*
- *carbon emissions*
- *small island developing states*
- *integrated resource*
- *non-polluting*
- *sustainable future*
- *power production*
- *environmental benefits*
- *transportation fuels*
- *photovoltaic*
- *electricity consumption*
- *onshore*
- *office of basic energy sciences*
- *hydroelectric dams*
- *thermal power*
- *farm operation*
- *greenhouse gas emissions*
- *wind events*
- *power electronics*
- *biomass fuels*
- *doe program*
- *rotor blades*
- *siting*
- *turbine blades*
- *electricity markets*
- *biogas plants*
- *climate control*
- *thermal systems*
- *solar cooker*
- *blades*
- *biogas*
- *fuel economy*
- *gasification*
- *heat pumps*
- *capital cost*
- *built environment*
- *biomass combustion*
- *biomass gasification*
- *wind velocity*
- *maldives*
- *roaring forties*
- *industrial technologies*
- *mitigation*
- *carbon sequestration*
- *digester gas*
- *ethanol*
- *ashrae*
- *waste heat*
- *biofuel*
- *producer gas*
- *environmental costs*
- *greenhouse gas abatement*
- *water turbine*
- *integrated assessment*
- *harmful emissions*
- *lead-acid batteries*
- *propulsion*
- *storage technology*
- *biodiesel*
- *ress*
- *tax incentives*
- *learning rates*
- *diesel*
- *environmental education*
- *vice president*
- *system voltage*
- *office of science*
- *market simulation*
- *republic of croatia*
- *long-term contracts*
- *national security*
- *fuel charge*
- *tax credit*
- *battery consortium*
- *public affairs*
- *tariff*
- *scheikunde*
- *domestic production*
- *utilization factor*
- *gasoline*
- *role of particles*
- *rice hulls*
- *american council*
- *warm springs*
- *financial incentives*
- *polluting*
- *insulation*
- *impact on the environment*
- *miles per hour*
- *economic costs*
- *dual fuel engine*

- *cost reduction*
- *metal recovery*
- *federal agencies*
- *local residents*
- *biological attack*
- *stand-alone*
- *land management*
- *storm damage*
- *taxpayer*
- *environmental protection agency*
- *life-cycle*
- *public benefits*
- *investor*
- *technology transfer*
- *contact dr*
- *economic information*
- *franchise tax*
- *co-operative*
- *virgin islands*
- *fall in price*
- *grounding*
- *groundwork*
- *option to purchase*
- *strait*
- *intermittent*
- *economic performance*
- *atmospheric sciences*
- *electroplating*
- *rinse water*
- *welding*
- *poultry litter*
- *data center*
- *economic dynamics*
- *world bank*
- *great plains*
- *turbulent*
- *open loop control system*
- *milieu*
- *gibraltar strait*
- *dinosaurs*
- *beryllium disease*
- *chronic beryllium disease*
- *tasmania*
- *climatology*

The results show that eigenvector centrality does an excellent job of identifying relevant terms. The highest-ranked term in the list is *renewable energy* itself. There is no mathematical reason that this must be the case; term centralities were calculated on a network that was built around *renewable energy*, but it need not be the single most central term in such a network. The fact that it is demonstrates the quality of eigenvector centrality as a metric for term relevance. We also see that all of the highly ranked terms in this list are in fact quite relevant. It is not until approximately the 80th term that we begin to see irrelevant terms such as *development mechanism*, *technological learning*, and *market barriers*. Furthermore, terms near the bottom of the ranking, with the possible exception of the final term, *climatology*, tend to be very irrelevant to *renewable energy*.

It is clear that eigenvector centrality is an effective metric for term relevance. Unfortunately, relevance is not our only concern. This metric seems to do a poor job of identifying the specific, little-known terms that are of interest to us in this project. *Solar photovoltaics* is perhaps the first relatively specific term encountered in this list, and it is found at the 41st position. For this reason, we find that eigenvector centrality is ineffective when used as a stand-alone term generation module. However, its relevance testing capabilities should not be overlooked, and it could potentially work well when used in conjunction with a filter that first removes all overly general terms.

3.3 Hit count extraction

Once a list of keywords related to the original seed term has been extracted, information is collected on the prevalence of each of these technologies over time. We seek a record of the level of activity in each field on a year-by-year basis. Since scientific publications are the predominant channel through which scientists and researchers publish their research results, the volume of related publications is a good indication of the popularity of a technology or term as a subject of research. Based on this intuition, we can hypothesize that the number of academic publications related to a particular topic is a good metric for its prevalence. This is clearly a very advantageous approximation since acquiring yearly publication counts is straightforward using publication search engines. We simply query a search engine with the seed term, narrowing the search to the particular year in which we are

interested, and extract the hit count from the returned results page. (The hit count is the number of publications from the given year containing the given term; nearly all search engines display this value on their search results page.)

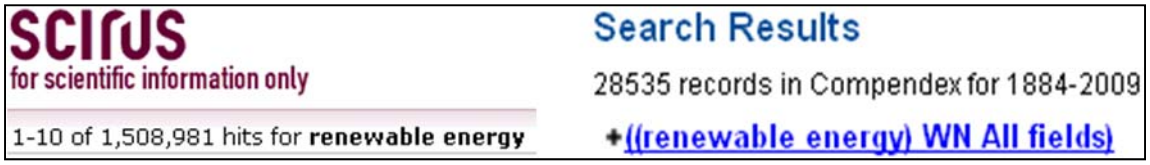

Figure 9. Hit counts for *renewable energy* given by Scirus on the left and Compendex on the right.

In determining the hit count for a particular term, we can ensure that the term is taken in its proper context by querying the search engine on both the term and the original seed term. This causes the search engine to only admit publications that include both terms, thus avoiding publications that use the term in a different context than the one in which we are interested. As an example, the term *corn* is relevant to energy technologies for its role in ethanol production. A Scirus search for *corn* returns approximately 2.1 million hits. A search for *corn* + *renewable energy*, on the other hand, returns only about 90,000 hits. The discrepancy is primarily due to the large number of articles about corn that are unrelated to energy. For our purposes, 90,000 is the more accurate count.

However, there are two potential downsides to requiring that the seed term be included in our queries. The first issue is that a term may represent an important component of the field, but is rarely mentioned in articles that also mention the seed term. Examples include technologies that have a wide range of applications, only one of which is related to the seed term. *Genetic engineering* is an example when *renewable energy* is the seed. Genetic engineering is an important component of biofuel development, so new genetic engineering research can indirectly affect the renewable energy field, often in ways which are unforeseen *a priori*. We would therefore be interested in articles about genetic engineering research for its own sake, even when they make no explicit mention of renewable energy. Only considering articles that contain both terms *genetic engineering* and *renewable energy* would result in a lower hit count, which understates the actual number of relevant articles.

There is a second downside to including the seed term in the query. In some cases, a term may appear in an article that is related to the seed term, but that never explicitly mentions the seed term because it is implicit from the context. An article about windmills, for example, is clearly relevant to renewable energy, but may never actually use the expression *renewable energy*. We would like this article to be included in the publication count for *windmills*, but requiring *renewable energy* to be included in the search query would prevent that from happening.

So on one hand, including the seed term in our search queries prevents artificially high hit counts that result from articles that use a term in an irrelevant context (e.g. *corn*). On the other hand, including the seed term can produce an artificially low hit count by not including relevant articles that discuss the term for its own sake, without discussing applications (e.g. *genetic engineering*), or by not including articles that do not mention the seed term because its relevance is obvious from context (e.g.

windmills). The choice of whether to include the seed term in the search query therefore involves a tradeoff, and is left as a customizable parameter in our software.

Just as we can use multiple sources for keyword extraction, we can also use one or more search engines to find a term's hit counts. Furthermore, the source(s) that we use for hit count extraction need not be the same as those used for keyword extraction. These are two independent processes; keywords can be extracted from whichever sources we choose, and their hit counts can be analyzed using whichever sources we choose. If multiple sources are used for hit count extraction, it is important to note that we do not have a good way to merge the hit counts we find from each source. Instead, each of the various sources' hit counts are kept separate, and each is thought of as a unique experiment. For example, the hit counts for *wind power* from 2006-2008 are (141,723, 836,044, 238,070) from Scirus, and (2,205, 2,982, 3,125) from Engineering Village's Compendex. We store these values separately and only compare Scirus hit counts with Scirus hit counts for other terms, and Compendex hit counts with Compendex hit counts for other terms.

In deciding what search engines to use for hit count extraction, we need to consider the size and focus of the underlying database, as well as the feasibility of using it for automated hit count collection. In general, the larger the database's collection of articles on our topic of interest, the more statistically significant its hit count trends will be. For feasibility, the database must allow thousands of queries to be submitted in a relatively short period of time, and it must also display the resulting hit counts in a manner that is extractable with software. With these considerations in mind, we are currently using Compendex and Scirus to gather hit counts, as they tend to give relatively large counts even for obscure energy-related terms. In contrast, we have decided that Inspec, the third database used in the term collection step, is too small of a database to be effective in this hit count collection step. Google Scholar is large enough to be a feasible choice, but it does not allow high-frequency querying and blocks our software if this is attempted.

To review the hit count extraction process, first, one or more search engines are chosen. If multiple search engines are used, each is treated as a separate experiment. A range of years over which to collect the hit counts is specified. Finally, it must be decided whether or not to include the seed term in each query. Our software then queries the search engines and extracts hit counts for every term in the list, over every year in the time range. These hit counts are then saved to disk for further analysis.

3.3.1 Software implementation and performance

For the implementation of hit count extraction and storage, we have a function called store_hit_counts, which takes six input parameters. These include the list of terms, an indication of whether to use Scirus or Engineering Village's Compendex, the list of years over which we wish to collect hit counts, the center term, or "False," if we do not wish to include it in our queries, the filename of a local database into which we write the results, and the number of terms whose hit counts we have already extracted in previous executions of the software, in case a previous execution was stopped prematurely, so that it can be re-run from where it left off.

For every term whose hit counts had not previously been extracted, as indicated by the term list and number of terms which had been previously analyzed, and for every year in which we are interested, the function queries the appropriate search engine for the hit count of that term in that year. The query and subsequent hit count extraction are executed using Python's URL library and Regular Expression modules. Just as with term collection, we construct the URL of the term's search results page, with or without the center term included in the query, as specified by the fourth input parameter. The contents of this page are read and regular expressions are used to locate and extract the hit count.

Results are saved to a local SQLite database file using Python's DB-API2 interface. Using an SQL database allows results to be easily and instantly accessible whenever needed. The database consists of a single table with two columns. The value in the first column is a term, and the value in the second column is a comma-separated list of that terms hit counts for each year. The years themselves are not included in the database, but can be indicated in its filename if we choose.

The runtime of the hit count collection module is proportional to the product of the number of terms and the number of years over which to collect hit counts. For each term and for each year, we must open and read a search results page to extract the hit count. The value must also be written to the local database, but this operation is nearly instantaneous. Extracting the hit counts for 10 terms over 3 years takes approximately 25 seconds, extracting the hit counts for 20 terms over 3 years takes approximately 45 seconds, and for 10 terms over 6 years, it takes approximately 45 seconds.

3.3.2 Results

Results are now presented for hit count extraction. Eight terms that represent broad areas within the renewable energy field are used, as well as *renewable energy* itself. Hit counts are extracted for each of these terms across the ten-year period from 1999 through 2008. Four different variants of the hit count extraction parameters are demonstrated: The Scirus search engine without including *renewable energy* in the search query, Scirus with *renewable energy* in the query, Compendex without including *renewable energy*, and Compendex with *renewable energy*.

Hit count extraction: Scirus, 1999-2008, *renewable energy* not in query:

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
wind	51804	67082	88498	113820	153390	201753	280560	438953	3574328	807572
solar	50109	100225	88727	110510	119025	152620	230390	315810	2455587	667056
hydro	10369	12401	16279	22837	31063	40475	53497	76242	562812	118440
geothermal	2698	3220	4922	6027	8571	11129	16171	24238	172650	53157
biofuel	258	256	444	663	1203	1511	2612	5862	90163	47116
nuclear	97909	115115	146110	174365	219403	268775	370554	477758	4318776	1070525
coal	17901	22756	33378	58047	60674	76286	103547	149920	1011542	295900
oil	104821	121553	160125	213522	258014	320429	416679	572632	4671434	1277368
renewable energy	3200	4603	7541	10763	21403	24265	34033	89392	714691	224122

Hit count extraction: Scirus, 1999-2008, *renewable energy* in query:

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
wind	1150	1667	3097	4139	7286	9976	14215	23799	335793	87420
solar	1529	2039	3560	4960	8356	11070	14532	54080	369381	95409
hydro	448	728	1231	1704	3089	4241	5869	9139	102368	19623
geothermal	415	651	1127	1469	2436	3233	4645	7329	50506	17274
biofuel	54	91	166	217	399	605	1012	2244	28622	11622
nuclear	1031	1308	2408	3088	4855	6215	9104	15183	173603	49115
coal	851	1427	2391	3273	5117	6986	9865	16409	99544	42874
oil	1222	1768	3186	4267	7186	9268	13317	22812	215474	76816
renewable energy	3200	4603	7541	10763	21403	24265	34033	89392	714691	224122

Hit count extraction: Compendex, 1999-2008, *renewable energy* not in query:

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
wind	2453	2903	3319	3892	4206	6188	6507	6269	6542	6610
solar	2853	3380	4174	5873	6057	7198	8340	8065	7875	8768
hydro	502	570	629	671	918	1075	1675	1750	1266	1359
geothermal	400	339	414	603	647	669	704	810	610	725
biofuel	26	23	28	41	46	90	128	152	168	273
nuclear	13451	14033	17409	18994	20683	25622	26027	25037	25668	23235
coal	1942	2410	2938	3629	3484	3941	4570	4607	3871	4151
oil	6130	6794	7126	8431	8717	11342	16072	13349	11859	14219
renewable energy	620	644	684	935	1281	1486	1709	2031	2336	4025

Hit count extraction: Compendex, 1999-2008, *renewable energy* in query:

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
wind	105	93	100	160	232	306	332	391	599	735
solar	191	192	183	306	363	396	437	466	549	807
hydro	22	14	22	22	47	57	63	67	58	85
geothermal	15	13	23	42	56	44	45	69	143	62
biofuel	0	2	2	2	5	11	26	18	17	47
nuclear	25	17	35	52	65	78	95	78	79	127
coal	16	17	33	37	58	76	91	132	85	175
oil	34	22	34	56	82	123	146	196	182	333
renewable energy	620	644	684	935	1281	1486	1709	2031	2336	4025

The results show that Scirus produces much higher hit counts than Compendex does. Large hit counts are helpful for ensuring the statistical significance of our growth rate calculations, so we typically use Scirus for hit count extraction. An interesting observation is that Scirus produces unusually large hit counts in the year 2007, for all terms. The reason for this is unknown, but since the effect is roughly the same across all terms, it is not an issue of concern.

The results also demonstrate the effect of including *renewable energy* in the search query. While this causes a significant reduction to all hit counts, it is not a particularly disruptive operation in other ways. For example, the rank ordering of all terms in a given year tends to remain roughly the same whether or not *renewable energy* is included in the query. Since it does not introduce anomalies into the hit counts, we prefer to include it in queries to prevent unnaturally large hit counts from terms being used out of context. And though the hit count reduction can decrease the statistical significance of our data, the counts remain large enough as long as we use Scirus.

3.4 Growth rate analysis

Once the hit counts for all terms have been extracted and saved, they can be used to calculate the respective growth rates. These, in turn, are used to rank the list of terms. As stated previously, we are most interested in terms with hit counts that are relatively low at present but that are quickly increasing, because these are likely to represent fields about which an expert would wish to be alerted.

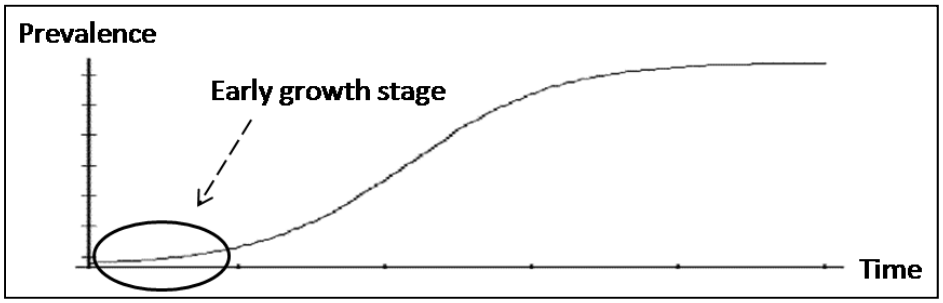

Figure 10. Growth rate analysis aims to identify technologies that are not yet well known but are nevertheless rapidly incerasing in prevalence.

There are many options for precisely how to define the growth rate of a term's hit counts. We consider four possibilities: the difference between the final year's hit count and the initial year's hit count, the ratio of the final hit count to the initial hit count, the rate of an exponential curve fit to the hit count time series, and the term's average year of publication. Each has its own advantages and disadvantages, described in this section.

Our first metric used for growth rate calculation is the numerical difference between the term's final and initial hit count. This straightforward measure represents the absolute amount by which the term's prevalence has grown over the entire time period. It can be thought of as the slope of a line passing through the initial and final points, multiplied by the length of the time period. Since all terms use the same time period, this multiplicative factor is constant across all terms. Therefore, the difference between final and initial points is essentially equivalent to the slope of the line between

them, as they each produce the same ranking of terms. Although this simple metric is intuitively appealing, it has two downsides. The first is that hit counts from all years besides the first and last are ignored. The second, and probably more important, downside is that terms with large initial hit counts are distinctly favored over terms with smaller initial hit counts. For example, consider the case where term A's hit counts increase from 1,000,000 in the initial year to 1,000,100 in the final year, and term B's hit counts increase from 1 in the initial year to 50 in the final year. Using the metric described above would lead us to conclude, erroneously, that term A is faster-growing than term B; in truth we would likely be more interested in term B, as its rate of growth is much higher than that of term A, which essentially stagnated.

We address the issue of the difference metric favoring already-prevalent terms with the ratio metric. Here we simply take the ratio of the final year's hit count to the initial year's hit count. Since this measure considers growth proportional to the initial number of hit counts, term B from the example above would be declared as faster growing than term A. Since $\log\frac{X}{Y} = \log X - \log Y$, the logarithm of the hit count ratio is equivalent to the difference in logarithms of the hit counts. And since the logarithm is a monotonic function, the logarithm of the ratio produces the same term rankings as the ratio itself. Therefore, this metric is equivalent to taking the difference between the logarithms of the final and initial hit counts, as opposed to the difference between the final and initial hit counts themselves, as in the previous method. Since the rate of change of $\log x$ decreases with increasing x, this method favors terms which begin with small hit counts. In this project, we are most interested in terms that are relatively unknown, but which are rapidly increasingly in prominence. Therefore, the ratio metric typically outperforms the difference metric.

Use of the ratio metric has one caveat: the growth rate is undefined if the initial year's hit count is zero. There are a few possible workarounds to this problem. One possibility is to replace the zero with the lowest nonzero hit count the term acquires in a future year. If the hit count remains zero until the final year, then we accept that its ratio is infinite, and rank it against other infinite-ratio terms based solely on the final year's hit count. A second possibility is to replace an initial zero with some small number, say 0.1. The choice of this number is unfortunately somewhat arbitrary, but the smaller it is defined to be, the higher the ranking a term with an initial hit count of zero will receive. We have chosen to use the second method, as the results it produces seem to coincide well with intuition.

The two metrics discussed above only consider the hit counts from the initial and final years of the range. Two other metrics are presented here, which consider all hit counts. The first involves fitting an exponential curve to the hit count time series and taking the exponential parameter as the growth rate. In other words, if the best-fit exponential curve to our data is of the form $a \times e^r$, then r would represent our growth metric. This metric corresponds with the intuition that technologies tend to grow at exponential rates.

Another possibility is the average year of publication. To calculate this growth indicator, we sum over all years y_i the product of the year and the hit count from that year h_i, and divide this result by the

sum of the hit counts across all years. Thus, $Average\ year\ of\ publication = \sum_i (y_i \times h_i) \Big/ \sum_i h_i$.

Intuitively, this metric is an indication of the time frame in which the term was the most prevalent. A later average year of publication indicates larger growth than an earlier average year of publication.

Thus far, we have discussed calculating growth rates based on a term's "raw" hit counts over time. This process can be made more robust by normalizing each year's hit count against the total number of articles in the publication database for that particular year. This prevents us, for example, from misinterpreting a change in a term's hit counts from year to year as a change in its prevalence when in reality it simply reflects a change in the number of articles indexed by the search engine.

This normalization is performed by first creating a time series of hit counts of a highly general term, which represents the total number of articles in the broad field about which we are interested. We could acquire a reasonably accurate estimate of the total number of articles indexed by the entire search engine by using the hit count of a term such as *the,* since almost every article is likely to contain this term. However, because we are interested in the prevalence of each term within our field of interest, it would make more sense to perform this normalization with respect to the size of the body of documents that are relevant to our field. We therefore use *renewable energy* as our normalizing term. *Energy* is another option, but we prefer to include *renewable* so as to avoid articles which use the term *energy* in different contexts (e.g. athletics). Once we have the hit count time series for our general term, normalized hit counts are determined by dividing all raw hit counts for a given year by the hit count of the general term for the same year.

As a concrete example, the Scirus hit counts for *wind power* and *solar power* over the years 2006-2008 are (15,784, 176,363, 58,278) and (10,930, 106,838, 49,146), respectively. We immediately notice the large spike in hit counts for each term in the year 2007. Without normalizing these hit counts, we might be led to believe that wind power and solar power were much more significant renewable energy technologies in 2007 than they were in 2006 or 2008. But the hit counts for *renewable energy* over the years 2006-2008 are (89,282, 714,662, 224,041), so we see that in fact the entire body of literature indexed by Scirus on renewable energy was unusually high in 2007. There are several possible explanations for this, but they are not important for our purposes; we simply normalize our terms' hit counts accordingly. After dividing each the hit counts for *wind power* and *solar power* by the corresponding years' hit counts for *renewable energy*, we have the following normalized hit count time series for *wind power* and *solar power*: (0.177, 0.247, 0.260) and (0.122, 0.149, 0.219), respectively. Given these normalized counts, we see that wind power and solar power did not, in fact, have unusually high prevalence in the renewable energy field in 2007, but instead slowly increased in prevalence over all three years.

Finally, we note that for the most part, we are only interested in the *ranking* of terms based on their growth rates, and not the actual magnitude of the rate for each term. In fact, each of our different growth rate algorithms produces rates with different physical units, e.g. the difference of final minus initial hit count has units of hit counts per year, whereas the ratio of final to initial hit count is dimensionless. For this reason, the magnitude of the growth rate for a single term is not very

informative. But the term ranking implied by the growth rates is quite informative, as it helps us identify the fastest-growing, most interesting terms. It is important to note that running any one of our growth rate algorithms on normalized hit counts can produce a different term ranking than running the same algorithm on the raw hit counts, so normalization can have a significant effect.

3.4.1 Software implementation and performance

Our growth rate software is executed by calling the function *get_ranked_growths*, which takes as its parameters a local database of hit counts, an indication of which of the available years from the database to consider, and an indication of which growth rate metric to use. The function queries the local database, reading, for each term, the hit counts over the years indicated in the input parameters. Each term's rate of growth is then calculated by applying the specified growth metric to the list of annual hit counts. Finally, the terms are sorted according to their growth rates to produce a ranked list.

Each growth metric is implemented as a function that takes a list of hit counts as input and produces a numerical growth value as output. This separates the process of querying the database for hit counts from the process of analyzing these hit counts to calculate a growth rate. This modularity makes it easy to incorporate new growth rate metrics into the software, as one simply needs to write a new function that produces a rate from a list of hit counts.

The runtime of this module depends on which growth rate metric is being used. For metrics that consider only hit counts for a constant number of years, such as the ratio or difference of the final and initial years, the runtime is proportional to the number of terms in the database. But for metrics that consider all (or a fraction of) the available years, the runtime is proportional to the product of the number of terms and the number of years. However, since the process consists only of querying a local database and performing arithmetic operations on the results, most operations complete almost instantly even for cases involving thousands of terms and a growth rate metric that incorporates hit counts from all years.

3.4.2 Results

Growth rate analysis results are presented here. We analyze the nine terms that were previously used in the hit count extraction demonstration. For simplicity, we arbitrarily choose to consider only the hit counts gathered from Scirus, with *renewable energy* included in the search query. These hit counts, first presented in section 3.3.2., are reproduced here:

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
wind	1150	1667	3097	4139	7286	9976	14215	23799	335793	87420
solar	1529	2039	3560	4960	8356	11070	14532	54080	369381	95409
hydro	448	728	1231	1704	3089	4241	5869	9139	102368	19623
geothermal	415	651	1127	1469	2436	3233	4645	7329	50506	17274
biofuel	54	91	166	217	399	605	1012	2244	28622	11622
nuclear	1031	1308	2408	3088	4855	6215	9104	15183	173603	49115
coal	851	1427	2391	3273	5117	6986	9865	16409	99544	42874
oil	1222	1768	3186	4267	7186	9268	13317	22812	215474	76816
renewable energy	3200	4603	7541	10763	21403	24265	34033	89392	714691	224122

Using these hit counts, growth rates are calculated and presented over two different time periods. We first display growth rates over the initial three years of our range, 1999 through 2001. Following this, we display growths over the entire range, from 1999 through 2008. This methodology provides a form of back testing and verification of our methods. We can imagine that it is currently the year 2002. Given knowledge of the past three years, we make predictions as to which terms will show growth over the course of the decade. These predictions can then be tested against the actual ten-year hit counts.

Ranked growth rates are displayed for all terms, over both the short-term and long-term time periods, using all four growth metrics, i.e. ratio of final hit count to initial hit count, difference of final hit count and initial hit count, exponential curve fitting rate, and average year of publication. Furthermore, each of these four metrics is applied with and without normalization of hit counts to *renewable energy*'s hit counts.

Growth rates, ratio, without normalization:

	1999-2001
biofuel	3.074
coal	2.810
hydro	2.748
geothermal	2.716
wind	2.693
oil	2.607
renewable energy	2.357
nuclear	2.336
solar	2.328

	1999-2008
biofuel	530.037
wind	291.994
solar	241.583
hydro	228.500
renewable energy	223.341
oil	176.329
nuclear	168.383
geothermal	121.701
coal	116.973

Growth rates, ratio, with normalization:

1999-2001		1999-2008	
biofuel	1.304	***biofuel***	2.373
coal	1.192	***wind***	1.307
hydro	1.166	***solar***	1.082
geothermal	1.152	***hydro***	1.023
wind	1.143	***renewable energy***	1.000
oil	1.106	***oil***	0.790
renewable energy	1.000	***nuclear***	0.754
nuclear	0.991	***geothermal***	0.545
solar	0.988	***coal***	0.524

Growth rates, difference, without normalization:

1999-2001		1999-2008	
renewable energy	4341	***renewable energy***	711491
solar	2031	***solar***	367852
oil	1964	***wind***	334643
wind	1947	***oil***	214252
coal	1540	***nuclear***	172572
nuclear	1377	***hydro***	101920
hydro	783	***coal***	98693
geothermal	712	***geothermal***	50091
biofuel	112	***biofuel***	28568

Growth rates, difference, with normalization:

1999-2001		1999-2008	
wind	0.051	***wind***	0.110
coal	0.051	***solar***	0.039
oil	0.041	***biofuel***	0.023
hydro	0.023	***hydro***	0.003
geothermal	0.020	***renewable energy***	0.000
biofuel	0.005	***geothermal***	-0.059
renewable energy	0.000	***nuclear***	-0.079
nuclear	-0.003	***oil***	-0.080
solar	-0.006	***coal***	-0.127

Growth rates, exponential rate, without normalization:

	1999-2001		1999-2008
biofuel	0.562	***biofuel***	0.656
coal	0.517	***solar***	0.590
hydro	0.505	***wind***	0.577
geothermal	0.500	***renewable energy***	0.573
wind	0.495	***hydro***	0.556
oil	0.479	***oil***	0.533
renewable energy	0.429	***nuclear***	0.520
nuclear	0.424	***geothermal***	0.502
solar	0.423	***coal***	0.499

Growth rates, exponential rate, with normalization:

	1999-2001		1999-2008
biofuel	0.133	***biofuel***	0.083
coal	0.088	***solar***	0.017
hydro	0.077	***wind***	0.004
geothermal	0.071	***renewable energy***	0.000
wind	0.067	***hydro***	-0.017
oil	0.051	***oil***	-0.039
renewable energy	0.000	***nuclear***	-0.052
nuclear	-0.004	***geothermal***	-0.071
solar	-0.006	***coal***	-0.073

Growth rates, average year of publication, without normalization:

	1999-2001		1999-2008
biofuel	2000.360	***biofuel***	2006.676
coal	2000.330	***wind***	2006.573
wind	2000.329	***solar***	2006.526
hydro	2000.325	***renewable energy***	2006.480
geothermal	2000.325	***hydro***	2006.452
oil	2000.318	***nuclear***	2006.452
nuclear	2000.290	***oil***	2006.395
solar	2000.285	***geothermal***	2006.192
renewable energy	2000.283	***coal***	2006.143

Growth rates, average year of publication, with normalization:

	1999-2001		1999-2008
biofuel	2000.088	***biofuel***	2003.593
coal	2000.057	***solar***	2003.129
hydro	2000.050	***wind***	2003.057
geothermal	2000.047	***renewable energy***	2003.000
wind	2000.045	***hydro***	2002.912
oil	2000.034	***oil***	2002.758
renewable energy	2000.000	***nuclear***	2002.666
nuclear	1999.997	***geothermal***	2002.594
solar	1999.996	***coal***	2002.591

Using *biofuel* from this example, it can be seen that the difference metric favors terms with large initial hit counts. For all three metrics besides difference, *biofuel* is rated as the fastest-growing term over both the early and full time period, whether the hit counts are normalized or not. The un-normalized difference metric, however, lists *biofuel* as the slowest grower over both time periods. The normalized difference metric is only slightly better and ranks *biofuel* near the middle for both periods. The difference metric artificially lowers the rank of terms with low initial hit counts, which runs contrary to our goals of identifying young, relatively unknown technologies.

Another interesting term from this example is *solar*. For the first few years of our time frame, *solar* grew relatively slowly. Its early-year ranking is very low with every metric, with the exception of the un-normalized difference metric, with which it is ranked second due to its initially high hit count. In the later years, however, *solar* underwent large growth, increasing its ranking on all metrics, except for the un-normalized difference, where it maintained the same high ranking. Using our back testing methodology, we see that only the un-normalized difference would have properly predicted this growth. This is not to say, however, that this metric should be used; it made the proper prediction because of its inaccurate assessment of the growth of *solar* in the early years.

Two terms whose long-term stagnancy would not have been successfully predicted are *coal* and *geothermal*. In the case of *coal*, no metric would have predicted its decreased growth in the later years, given its large growth initially. In the case of *geothermal*, only the un-normalized difference would have predicted its slowdown, but this is mostly due to an unrealistically low growth ranking in the early years.

Overall, we see that the ratio, exponential rate, and average year of publication metrics all produce similar results, and best meet our interests of identifying technologies at their early stages of growth. We also see that among these three metrics normalization has only a slight effect on the growth rate rankings. (This is not true of the difference metric, on which normalization has a significant impact.) In this project, we typically choose to use the normalized ratio and normalized exponential rate metrics.

3.5 Keyword extraction and analysis wrap-up

This chapter has described the process of keyword extraction and analysis. We begin with a seed term that represents our general field of interest. In our case, this term is *renewable energy*. That term is used to automatically gather a large number of other, related, terms. The yearly publication counts are then identified for each of these terms over a period of time in which we are interested. Finally, these publication counts are used to calculate a growth rate for each term and to rank them accordingly. The terms that are ranked highly can then be presented to an expert as potentially important technologies to monitor in the near future.

The next chapter presents a technique known as Latent Semantic Analysis for identifying relationships among a set of terms. The process is computationally intensive, with a runtime proportional to the square of the number of terms given as input. For this reason, it is important that we first reduce the list of terms to a relatively small number of interesting ones, using the growth rate techniques already described in this chapter.

Chapter 4: Latent Semantic Analysis

4.1 Overview

In previous chapters, we have discussed representing an entire research field as a single keyword. That keyword's hit counts for a given year are taken as a representation of the prevalence of the field at that time. This idea works well as a first approximation. However, there are two distinct problems with representing a field by a single term. The first is that the term may have synonyms, or other terms that convey the same or similar meaning. The second is that the term may be a homonym, having multiple meanings. In either case, the hit count of the single term will be an inaccurate representation of the prevalence of the field as a whole.

In the case of synonymy, searching for a single term produces a hit count that is artificially low. Consider the case of *oil* and *petroleum*. Some documents use the term *oil*, while others use *petroleum*, but all of these documents are inherently discussing the same topic. We are interested in the prevalence of the topic, rather than of either of the specific terms used to describe it. From the year 2008, Scirus finds 1,277,277 documents containing *oil* and 305,562 documents containing *petroleum*. But in a search for *oil OR petroleum*, Scirus finds 1,448,608 documents. This last hit count is the most representative of the field's prevalence.

A term need not have exact synonyms, as in the case with *oil* and *petroleum*, to be under-representative of the field as a whole. For example, if we are interested in the prevalence of renewable energy, we would include articles in the hit count that contain *wind power* or *solar power*, even if those articles do not also explicitly include the term *renewable energy*. So to avoid artificially low hit counts, we can include all terms that are either exact synonyms or related terms to our search keyword.

The opposite problem can occur if the term is a homonym, having multiple meanings or interpretations. In this case, its hit count would be artificially high. Consider, again, the renewable energy field. A search for the term *energy* would return many articles of interest, but it would also include some irrelevant articles that use *energy* in a different sense, such as the effectiveness of coffee for an energy boost. To restrict the articles to relevant ones, and to thereby reduce the hit count and make it more representative of the field's prevalence, an *AND* can be included in the search query along with a term that is a synonym of the search term when considered in the appropriate context. We can search for *energy AND power*, for example, rather than just *energy*. (While energy and power are not technically synonyms, it is a good assumption that many energy-related articles will also include the term *power*.)

We need a tool to automatically identify relations between terms, so that our search queries can be modified to mitigate the hit count inaccuracies caused by synonyms and homonyms. One possibility is a technique known as clustering, which consists of grouping keywords into a number of disjoint clusters of related terms. The clustering algorithm uses co-occurrence of terms in the documents indexed by a search engine to determine which terms are grouped with which other terms. If the algorithm finds, for example, that *solar* and *sun* frequently appear in the same documents, but

that *wind* tends to not appear with either term, *solar* and *sun* will be put in a separate cluster from *wind.* This technique can be effective for dealing with synonymous terms. When interested in the prevalence of a particular field, we would simply include all the terms in its cluster in the search query.

Unfortunately, clustering is incapable of properly representing homonyms. Since every term is contained in exactly one cluster, all terms are inherently given a single sense or meaning. For example, if we were to form two clusters from the terms *power*, *coffee*, and *energy*, we would likely see one cluster containing *power* and the other containing *coffee*, while *energy* would be contained in one or the other, but not both. Thus, clustering does not allow for the possibility of multiple senses or meanings of a term.

Latent Semantic Analysis, or LSA, is another technique for identifying relationships between terms. It too uses term co-occurrence in documents to uncover these relationships. However, it presents a significant theoretical improvement over clustering, because it can accurately represent homonyms by not restricting a term to a single set of related terms. From a set of terms, LSA produces a set of so-called "concepts," each of which is a weighted combination of *every* term in the set. A term that does not belong to a particular concept is still included, but is given a negligible weighting. A term with multiple meanings may be given a relatively high weighting in multiple concepts, demonstrating the fact that it belongs with more than one set of related keywords. This is the key characteristic that distinguishes LSA from clustering – a term can be a significant component of more than one concept.

LSA always produces a number of concepts equal to the number of terms given as input. If we provide *n* terms, LSA will produce *n* concepts that are each a weighted combination of all *n* terms. Just as the individual terms that make up a concept are weighted, so are the concepts as a whole. In some sense, the concepts are weighted according to their prevalence in the documents indexed by the search engine. Loosely put, if a particular weighted combination of the terms is well-represented as a coherent concept by the documents in the search engine, LSA will give that concept a high weighting. (More technically, the weighting represents the variance in the prevalence, as opposed to the prevalence itself, of that concept across the documents in the search engine. This is discussed in more detail in the next section.) Since LSA weights the resultant concepts, those with low weights can be ignored. Furthermore, within the concepts that are kept, individual terms with low weights can be ignored. Thus, using LSA, a set of terms can be converted into a smaller set of concepts, each of which consists of a combination of a small number of the terms.

4.2 Mathematical foundations

LSA transforms a set of terms into a set of concepts using a rigorous statistical method. In this section, we describe this technique and how LSA applies it to bibliometric data. We first describe the well-established basic algorithm, and we then discuss our novel modifications.

4.2.1 Known algorithm

LSA is an application of a general statistical technique known as Principal Component Analysis, or PCA. PCA is used to transform a set of possibly correlated variables into a set of uncorrelated

variables. Each of the transformed variables is some linear combination of all of the original variables. Furthermore, for each transformed variable, PCA shows the degree to which the data vary along that dimension. This allows us to reduce the dimensionality of the data by ignoring variables along which the data have low variance. In LSA, the original variables are the occurrence counts of each term in a document, and the sample of data points is the set of documents indexed by our search engine. The uncorrelated, transformed variables are combinations of terms which we call concepts. After removing the concepts along which the data show low variance, we have reduced a set of n possibly correlated terms to a set of fewer than n uncorrelated concepts.

We first discuss PCA, the technique from which LSA is derived. Consider a simple example. We are measuring people's heights, first in feet, and then in inches. Both measurements are taken independently and include some amount of measurement error. This gives us two variables – the number of feet, f, and the number of inches, i. For every person in our sample, we have a data point (f, i). Due to measurement error, we may see points such as the following: (5.1, 60.1), (5.51, 65.5), (6.1, 72.5). Each point can be plotted on a graph with f on the x-axis and i on the y-axis. Of course we will find that all the points lie close to the line which passes through the origin and has a slope of 12 inches/foot.

Given this set of data, it is clear that f and i are highly correlated. Given a measurement in feet, a relatively accurate estimate of the measurement in inches can be made, and vice versa. But if we use PCA on this data, it will produce two combinations of these variables which are in fact completely *un*correlated. The two combinations are $f + 12i$ and $12f - i$. It is easy to see why these two combinations are uncorrelated. $12f - i$ is approximately 0 for all data points, and does not vary in any systematic way with the value of either f or i. Therefore knowledge of $12f - i$ does not provide any information about $f + 12i$. Note that by this reasoning, $12f - i$ is actually uncorrelated with any other combination of f and i. However, PCA imposes the additional constraint that the combinations must be orthogonal to one another. If $12f - i$ is represented as the vector $12\hat{f} - \hat{\imath}$, the only vector orthogonal to it in (f, i) space is $\hat{f} + 12\hat{\imath}$, and hence PCA chooses $f + 12i$ as the variable uncorrelated with $12f - i$. By imposing both an uncorrelatedness constraint and an orthogonality constraint, PCA can be thought of as a coordinate rotation from axes which were orthogonal but correlated to axes which are orthogonal and uncorrelated.

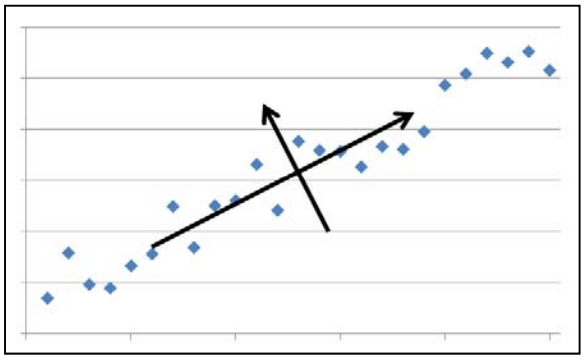

Figure 11. From this two-dimensional data set, PCA identifies two orthogonal vectors, one of which (the "principal" component) is in the direction of maximum variance in the data.

We have described one of the two functions of PCA; namely, to identify a set of orthogonal axes which are uncorrelated with one another. These are the "components" from the term Principal Component Analysis. The second function is to identify how closely each one of these components represents the data; or to identify their "principality." This is done by projecting each data point onto the new set of components and calculating the variance of the data along each of the components. To continue with our example from above, consider the data point (5.1, 61). Its projection onto the dimension $\hat{f} + 12\hat{\imath}$ is 61.2. (The projection of a vector A onto a vector B is defined as $\frac{A \cdot B}{|B|}$.) Its projection onto the orthogonal direction, $12\hat{f} - \hat{\imath}$, is 0.017. If we have a different data point (6.1, 72), we find that its projection onto the dimension $\hat{f} + 12\hat{\imath}$ is 72.3, and its projection onto $12\hat{f} - \hat{\imath}$ is 0.100. We see from these two points that there is much higher variability in the projections along $\hat{f} + 12\hat{\imath}$ than in the projections along the orthogonal component $12\hat{f} - \hat{\imath}$. Thus PCA would identify $\hat{f} + 12\hat{\imath}$ as a more important component. This fits with the qualitative observation that the data approximately lie along the direction of $\hat{f} + 12\hat{\imath}$. Finally, the real value of PCA comes from dimension reduction. Since the projection of our data onto the components results in variability in one dimension but not the other, we can choose to ignore the invariant component and describe our data only in terms of the projection onto the principal component. This results in a more compact, informative data set, with the cost of a small loss of information. In our example, the dimensionality of the dataset is reduced from two different height measurements to one.

The implementation of PCA is fairly straightforward. We begin with a data matrix. Each column of this matrix represents a data point or experiment, and each row represents the value of a variable after mean-centering. Mean-centering is defined such that for a given variable X, if its observed mean across all values X_i is $\bar{X}$, then the mean-centered values that go into our matrix are $\overline{X_i} = X_i - \bar{X}$. In our height example, each column would represent a particular person, and would contain the mean-centered height in feet in the first row and the mean-centered height in inches in the second row.

Given a data matrix, we then perform the Singular Value Decomposition, or SVD, of this matrix. In general, the SVD factors an $m \times n$ matrix A into the form $U\sum V^T$. U is $m \times m$, and its columns are the eigenvectors of AA^T. $\sum$ is an $m \times n$ diagonal matrix that contains the square roots of the eigenvalues of AA^T along its diagonal. V^T is $n \times n$, and its columns are the eigenvectors of A^TA. The values in $\sum$, known as the "singular values" of A, are arranged in decreasing order down the diagonal of $\sum$. Each of these values $\sum_i$ corresponds to the column U_i of U. It is the columns of U that represent the transformed variables PCA identifies, and it is the associated singular values from $\sum$ that represent the degree to which the data fit each of these transformed variables. In our height example, the columns of U would be [1, 12] and [12, -1] to represent the transformed variables $\hat{f} + 12\hat{\imath}$ and $12\hat{f} - \hat{\imath}$, respectively. Correspondingly, the first singular value in $\sum$ would be quite large, representing the large variance of data along the $\hat{f} + 12\hat{\imath}$ direction, whereas the second singular value would be small, representing the relative invariance of the data along the $12\hat{f} - \hat{\imath}$ direction.

Thus far, we have discussed the intuition and mathematics behind PCA. LSA is a straightforward application of this general technique. We briefly discuss it in this section, and for a more in-depth introduction to LSA, the reader is referred to [Berry et al., 1995] and [Landauer et al., 1998]. Each of the

documents indexed by our publication search engine is treated as an experiment. For a given experiment (document), the variables are the occurrence frequency of each of the terms in that document. The occurrence frequency can be defined in one of several ways. The most straightforward is to simply use the count of the number of times the term appears in the document. A second definition of term frequency normalizes the count by the total number of words in the document. For example, if a term appears 5 times in a 500-word document, the occurrence frequency of that term in that document would be $5/500 = 0.01$. Finally, a third definition of term frequency, known as Term Frequency-Inverse Document Frequency (TF-IDF), further normalizes this quantity by the fraction of total documents that contain the term at least once. For example, if the term appears 5 times in a 500-word document, and appears at least once in 5% of all documents, then its TF-IDF value would be $\frac{5/500}{0.05} = 0.2$. Any of these metrics are valid frequency variables, but TF-IDF tends to be the most commonly used. However, using our modifications to the standard LSA algorithm (discussed in the next section), the issue of which term frequency metric to use becomes irrelevant.

Regardless of the definition we choose for term frequency, these data are then put into the PCA data matrix, which is named a term-document matrix when used for LSA. The rows of the matrix are all the terms, the columns of the matrix are all the documents, and the values $A_{i,j}$ in the matrix represent the frequency of term i's occurrence in document j. SVD of this term-document matrix is then performed, which gives uncorrelated combinations of terms in the columns of U. These weighted combinations of terms are the groupings we call "concepts." Furthermore, the singular values from ∑ tell us which concepts are the most representative of the data. Using this information, the initial list of (largely correlated) terms can be reduced to a smaller list of important and uncorrelated concepts.

4.2.2 Modifications to known algorithm

For our purposes, we use a novel modification to the standard LSA algorithm. The sheer number of documents to be searched renders the first step of standard LSA – generating the term-document matrix – highly impractical. Such a matrix could potentially have millions of columns. Additionally, determining each value in the matrix would require searching the complete text of each document to count term occurrences. Creating this matrix would only begin to be feasible if all documents were available on a local machine, which is an unacceptable limitation for our purposes given the sources that we are using (e.g., Scirus, Compendex), which are not under our control or available for local usage.

Our approach instead allows us to easily use the vast body of published literature available on the Internet as our document set. We only have to be willing to accept a simplifying assumption about our documents. The key assumption is that any given term occurs in any given document either 0 times or 1 time, i.e. that all document vectors in the term-document matrix are binary encoded and cannot contain values besides 0 or 1. While this is clearly not completely accurate, the assumption gives us the ability to use enormous online collections of documents, making this approach far superior to standard LSA for our purposes.

Instead of counting term occurrences in documents, we use only the hit count returned from a search engine to generate a term *covariance matrix*. If we are analyzing m terms, the covariance matrix A is $m \times m$, with each value $A_{i,j}$ representing the covariance between terms i and j. Such a matrix could have been derived from the term-document matrix, but for our purposes, our simplifying assumption allows us to construct an approximation of the covariance matrix from hit counts alone.

The derivation of the covariance matrix is as follows. We treat the numbers of occurrences of a particular term in each document as realizations of a random variable. Then, by definition, the covariance of two of these random variables (the covariance between two terms i and j in a term-document matrix A) is

$$cov(i,j) = E[(A_i - E[A_i]) \times (A_j - E[A_j])]$$

The expectations of A_i and A_j are calculated by taking the average of each value across all n documents for the given term. For example, $E[A_i]$, the expectation of the number of occurrences of term i, is $\frac{1}{n}\sum_{k=1}^{n} A_{i,k}$.

The outer expectation is similarly calculated by averaging the values for terms i and j across all n documents. Then our covariance expression becomes

$$cov(i,j) = \frac{1}{n}\sum_{k=1}^{n}\left(\left(A_{i,k} - \frac{1}{n}\sum_{l=1}^{n} A_{i,l}\right) \times \left(A_{j,k} - \frac{1}{n}\sum_{l=1}^{n} A_{j,l}\right)\right).$$

This expands to

$$cov(i,j) = \frac{1}{n}\left(\sum_{k=1}^{n} A_{i,k}A_{j,k} - \frac{1}{n}\sum_{l=1}^{n} A_{i,l}\sum_{k=1}^{n} A_{j,k} - \frac{1}{n}\sum_{l=1}^{n} A_{j,l}\sum_{k=1}^{n} A_{i,k} + \frac{1}{n}\sum_{l=1}^{n} A_{i,l}\sum_{l=1}^{n} A_{j,l}\right).$$

Combining the last three terms gives

$$cov(i,j) = \frac{1}{n}\left(\sum_{k=1}^{n} A_{i,k}A_{j,k} - \frac{1}{n}\sum_{k=1}^{n} A_{i,k}\sum_{k=1}^{n} A_{j,k}\right).$$

Simplification of this expression requires us to utilize our assumption that each value in the term-document matrix is either 0 or 1. With this simplification in mind, $\sum_{k=1}^{n} A_{i,k}$ represents the number of documents that contain term i, $\sum_{k=1}^{n} A_{j,k}$ represents the number of documents that contain term j, and $\sum_{k=1}^{n} A_{i,k}A_{j,k}$ represents the number of documents that contain both term i and term j. This is represented with this final covariance expression.

$$cov(i,j) = \frac{1}{n}\left(h_{i,j} - \frac{1}{n}h_i h_j\right).$$

In the above expression, $h_{i,j}$ represents the number of hits returned from a search for both terms, while h_i and h_j represent the number of hits returned for terms i and j respectively. n represents the number of documents being searched. We approximate n with the number of hits returned from a search for a large term that subsumes terms i and j. For our purposes, we use the field that is the focus of our case study, renewable energy, as the search term to acquire n.

Using our covariance expression, which was approximated using only hit counts under the assumption that a document can contain a given term zero times or one time, we can construct an approximate term-by-term covariance matrix. The exact term-term covariance matrix C of a term-document matrix A is defined as $C = AA^T$. (This properly equates each covariance $C_{i,j}$ to the dot product of term i's occurrence counts with term j's occurrence counts: $C_{i,j} = A_i \cdot A_j$.) We then make the assumption that our approximate covariance matrix is in fact exactly equal to AA^T. Recall from the original LSA algorithm that our concepts are defined as the columns of U from the SVD of A, and that the columns of U are defined as the eigenvectors of AA^T. Therefore, we wrap up our LSA approximation by defining our concepts as the eigenvectors of the hit-count-approximated covariance matrix. Due to the inaccessibility of a term-document matrix, this approximation is what we use in actual implementation.

The LSA calculation process, using our covariance approach, is illustrated in the figure below.

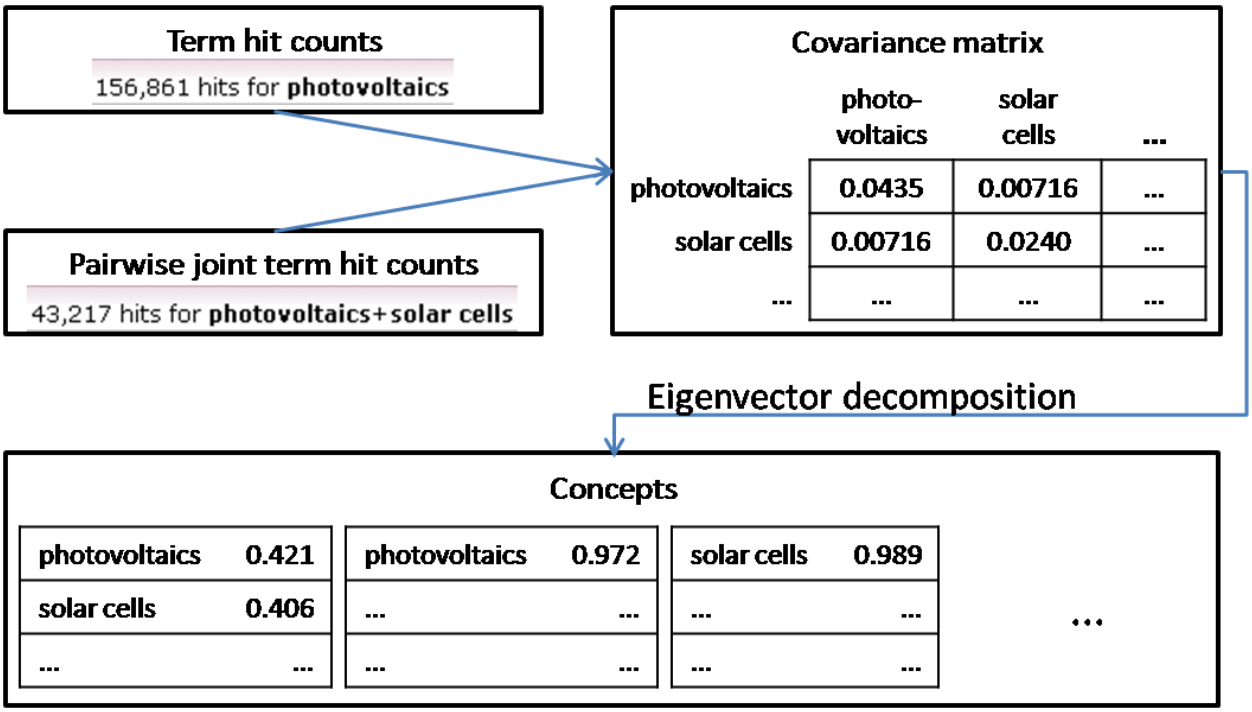

Figure 12. LSA calculation. Hit counts for each term and co-hit counts for each pair of terms are used to produce a covariance matrix. The eigenvectors of this matrix represent the concepts.

4.3 Software implementation and performance

To run our LSA software, the function *get_hit_counts* is called, which takes four parameters: the list of terms, the seed term, the year from which we want publication counts, and the publication search engine from which to extract these counts. This function then returns two objects. The first is the hit count of the seed term, using the year and search engine the user has provided. The second is an object that contains the joint hit counts of every pair of terms in our list. For example, if *solar power* and *wind power* were terms in our list, one entry in this object would contain the hit count for the query *solar*

power + wind power. There would also be entries for *solar power + solar power* and *wind power + wind power*, which are equivalent to the individual queries *solar power* and *wind power*, respectively.

Once we have run *get_hit_counts*, the function *lsa* is called, which takes as inputs the list of terms, the center term, and the center hit count and joint hit counts acquired from *get_hit_counts*. The function first creates a covariance matrix from the hit counts using the formula derived above, $cov(i,j) = \frac{1}{n}\left(h_{i,j} - \frac{1}{n}h_i h_j\right)$. Next, it uses Python's linear algebra package to calculate the eigenvalues and eigenvectors of this matrix. Finally, it rearranges the eigenvalues and eigenvectors into a form that sorts the terms within each concept in descending order of absolute weight, and sorts the concepts in descending order of associated eigenvalue.

Our LSA software runs in time approximately proportional to the square of the number of terms we are analyzing. The bottleneck in performance is joint hit count extraction. Given n terms, hit counts must be gathered for $\frac{n^2+n}{2} = O(n^2)$ queries. Each query requires reading and parsing a search results page, so this takes a non-negligible amount of time. The other potentially time-intensive step is the eigenvector calculation of the covariance matrix. As discussed in section 3.2.2.3.1, eigenvector calculation has an $O(n^3)$ runtime, unless memory swaps are required. When used for LSA, n tends to be relatively small, typically less than 100. With n this small, the runtime of the eigenvector calculation is negligible compared to the runtime of the joint hit count collection. So for the number of terms we typically use, the entire process has a runtime that is approximately $O(n^2)$.

When using 10 terms, our LSA software requires approximately 30 seconds. With 30 terms, it takes approximately 255 seconds. With 100 terms, it takes approximately 1 hour.

4.4 Results

LSA results are presented in this section, as applied to a set of nine renewable energy-related terms. The nine concepts are displayed in decreasing order of associated eigenvalue, so that those concepts that are displayed first are the most representative of the data. Within each concept, the term weights (equivalently, the eigenvector components) are sorted by descending absolute value, so that the terms displayed near the top of a concept are the most important terms to that concept.

LSA concepts:

Eigenvalue	67015
-------	
wind	0.8310
wind power	0.3680
turbine	0.2108
solar power	0.2056
geothermal	0.1998
coal	0.1940
fossil fuel	0.0938
solar cell	0.0222
heat pump	0.0205

Eigenvalue	15207
-------	
turbine	0.7136
solar power	-0.6000
geothermal	0.2858
wind	-0.1288
coal	-0.1041
fossil fuel	0.0973
wind power	0.0938
solar cell	-0.0515
heat pump	0.0256

Eigenvalue	9625
-------	
wind power	-0.5775
turbine	0.5291
solar power	0.4860
fossil fuel	-0.3532
solar cell	0.1435
geothermal	0.0511
heat pump	0.0324
coal	0.0321
wind	0.0167

Eigenvalue	35081
-------	
coal	0.9388
wind	-0.2656
fossil fuel	0.1855
geothermal	0.1085
turbine	0.0293
wind power	-0.0238
heat pump	0.0199
solar power	0.0075
solar cell	0.0053

Eigenvalue	13833
-------	
geothermal	0.8489
fossil fuel	0.2988
turbine	-0.2722
wind power	-0.2474
coal	-0.1727
solar power	0.1326
heat pump	0.0637
wind	-0.0539
solar cell	0.0230

Eigenvalue	3516
-------	
solar cell	0.9836
solar power	-0.1442
wind power	0.0810
turbine	-0.0621
fossil fuel	-0.0252
heat pump	-0.0202
wind	-0.0099
geothermal	0.0095
coal	0.0011

Eigenvalue	19647
-------	
wind power	0.6194
solar power	0.5511
wind	-0.4641
turbine	0.2356
coal	-0.1554
fossil fuel	0.0943
geothermal	0.0714
solar cell	0.0428
heat pump	0.0423

Eigenvalue	11225
-------	
fossil fuel	0.8495
geothermal	-0.3668
wind power	-0.2651
turbine	0.1742
solar power	0.1320
coal	-0.1231
solar cell	0.0789
wind	0.0588
heat pump	0.0313

Eigenvalue	2196
-------	
heat pump	0.9951
geothermal	-0.0610
fossil fuel	-0.0470
solar power	-0.0437
turbine	-0.0398
wind	0.0121
solar cell	0.0102
wind power	0.0088
coal	0.0004

The first concept appears to represent wind power generation, consisting largely of the terms *wind*, *wind power*, and *turbine*.

The second concept seems to represent various sources of energy, with the highly weighted terms *coal*, *wind*, *fossil fuel*, and *geothermal*.

In the third concept, the four highest-weighted terms are the same as those from the first concept, but ordered differently. This is a phenomenon that commonly occurs with LSA. Since n eigenvectors are produced that presumably represent fewer than n actual underlying concepts, some vectors appear as slightly rearranged versions of others.

The next four concepts are difficult to interpret. The highly weighted terms in each of these concepts are not related beyond the obvious fact that they are renewable energy-related. We see that the fifth and sixth concepts are another example of a pair that contain the same highly weighted terms, but ranked differently. This indicates that the set of terms *geothermal*, *fossil fuel*, *turbine*, and *wind power* are a well-represented group in the set of documents analyzed by our LSA algorithm.

When the results become difficult to interpret, as we see here, it is important to keep in mind that LSA concepts are simply a representation of the data given to it. These concepts are, mathematically speaking, components of the term-document data. They are not, however, *principal* components; these particular concepts are the fourth-ranked through seventh-ranked concepts out of nine, so it is perhaps not surprising that their results do not correspond with intuition as strongly as the first concept does.

Finally, of the least-principle concepts, we see that concept eight represents both solar power and wind power generation, with the terms *solar cell*, *solar power*, *wind power*, and *turbine*. Concept nine primarily represents geothermal power generation, containing *heat pump* and *geothermal* as its highest-weighted terms. However, its next two terms, *fossil fuel* and *solar power*, do not intuitively belong.

Given the terms used in this example, LSA have been would be expected to form concepts representing the following groupings: *wind*, *wind power*, and *turbine*; *solar cell* and *solar power*; *geothermal* and *heat pump*; and *coal* and *fossil fuel*. The results show that the first concept represents *wind*, *wind power*, and *turbine*, the eighth concept represents *solar cell* and *solar power*, and the ninth concept represents *geothermal* and *heat pump*. No concept directly represents the grouping of *coal* and *fossil fuel*, but the second concept approximates this with *coal* as its highest-weighted term and *fossil fuel* as its third highest-weighted term.

4.5 Data cleaning

LSA produces a set of concepts from a set of terms. The calculations that identify the concepts depend on the co-occurrences of every term with every other term (by way of translation into a term-term covariance matrix, whose eigenvectors represent the concepts.) This means that each individual term's presence can have a significant impact on the LSA concepts. If every term given as input to LSA

were satisfactorily relevant and specific to our topic, this would not be a problem. Unfortunately, due to our requirement of fully automated means of term collection, this is not always the case. We frequently encounter terms that are either overly general or irrelevant to our topic. And while in theory, LSA will still reveal term interrelationships even when the list contains these "noise" terms, such terms tend to obscure the resultant concepts, making them difficult to interpret.

For this reason, we seek methods to identify and remove offending terms from our list, before we provide it to LSA as input. A manual scan of the term list by a field expert would be one possible method for data cleaning. However, just as with term collection, we are only interested in fully automated means. We propose two such methods in this section, both of which are similar in structure. They rely on an iterative process which first calculates the concept vectors from the term list, then uses a heuristic to identify the most offending terms from these concepts, and finally deletes these terms from the list and repeats the process.

Our two methods differ in the heuristic used to identify which terms should be removed at each iteration. The first uses the calculated concept vectors to identify the most and least general terms, and removes them. The second identifies which terms' presence has the largest and smallest impact on the structure of the resultant concept vectors, and removes those terms. The first method is described in section 4.5.1, and the second method is described in section 4.5.2.

4.5.1 Term removal by level of generality

Using this method, we attempt to identify and remove terms which are either overly general or irrelevant to our topic. For example, if we are using renewable energy terms, our list might contain such overly general terms as *energy technologies* or *renewable energy* itself. The list also might contain irrelevant terms such as *education* or *impact*. The identification and removal of terms such as these is important to help maintain the focus and clarity of the LSA concepts derived from the terms we do not remove.

We use a relatively straightforward heuristic to identify general and irrelevant terms. First, LSA is run on the full list of n terms. This produces a set of n concept vectors, each of which is a weighted combination of all n terms. Then for each term, its level of generality is estimated by summing the absolute value of that term's weighting over every concept vector. Intuitively, general terms such as *renewable energy* will be important, highly weighted components of several concepts, whereas irrelevant terms such as *education* will have minimal weighting in most concepts. In cases where the term's weighting in a particular concept is negative, we ignore the sign and consider only the magnitude of the weighting. This is because a strong negative correlation between that term and others terms shows as much about the term's importance to the concept as a strong positive correlation would. Once the absolute values of weightings for each term have been summed as an indicator of generality, those with the highest and lowest weightings are removed from the list. LSA is then run on the reduced set of terms, giving a new set of concepts. Using these concepts, the new most and least general terms can be identified and removed. This iterative process can continue as many times as desired.

This algorithm for term removal by level of generality has three adjustable parameters. We can choose the number of times to iterate, the number of general terms to remove at each iteration, and the number of irrelevant terms to remove at each iteration. In general, the algorithm produces the most optimal results if several iterations are performed, with the removal of few terms at each iteration. However, each iteration is time-consuming, and in some cases it may be worthwhile to execute fewer iterations and remove more terms at each iteration, while sacrificing a small amount of optimality.

If runtime were not an issue, we would remove as few terms as possible at each iteration, but there would still be a choice to be made: we can choose to remove only the most general term, only the most irrelevant term, or both. This is a matter of preference, and the proper choice may vary depending on the types of terms we see in the term list. This is why we leave it to the user to decide how many general terms and how many irrelevant terms to remove at each iteration.

4.5.1.1 Software implementation and performance

For term removal by generality, the function *get_hit_counts* is called to extract all joint hit counts. The function *remove_general_terms* is then called, which, like the *lsa* function, takes in parameters for a term list, the seed term, and the hit counts as provided by *get_hit_counts*. But *remove_general_terms* also takes three additional parameters: the number of iterations of data cleaning to execute, and the number of general and irrelevant terms to remove at each iteration.

At each iteration, the function generates a covariance matrix from the joint hit counts and calculates the eigenvectors of this matrix. For each term, the absolute sum of its weights across all eigenvectors is then calculated, and the appropriate number of high- and low-weighted terms is removed from the list. The function *get_hit_counts* only needs to be executed before the first iteration; after this, each successive covariance matrix is constructed by referencing the appropriate subset of the original joint hit counts.

As with the *lsa* function, the performance bottleneck is in joint hit count extraction, which requires parsing $O(n^2)$ search results pages. Once the hit counts have been extracted, the remaining runtime is $O(i \times n^3)$ for i iterations, since for each iteration, eigenvectors of a matrix of size $O(n)$ are calculated, which requires $O(n^3)$ runtime. However, as with the *lsa* function, even though the analysis portion of *remove_general_terms* is asymptotically slower than the hit count collection portion, the relatively small size of n that we are using causes this step to run nearly instantaneously. When using 100 terms, for example, after all joint hit counts have been collected, 10 iterations of single-term removal requires only approximately 5 seconds.

4.5.1.2 Results

Results are now presented for the process of removing terms based on level of generality. We begin with a set of twenty renewable energy-related terms. We perform an initial LSA on these terms and display the results. For space considerations, only the four highest-weighted terms from the four highest-weighted concepts are displayed. Five iterations of data cleaning are then performed, eliminating the most and least general term from the list at each iteration. Finally, we show the results

of LSA performed on this reduced set of terms, again displaying only the four highest-weighted terms from the four highest-weighted concepts.

Initial LSA concepts:

Eigenvalue	29529
-------	
biomass	0.7743
co2	0.5115
biofuel	0.2822
developing countries	0.1856

Eigenvalue	10746
-------	
biofuel	0.5899
developing countries	0.5060
carbon dioxide emissions	0.4783
co2	-0.3561

Eigenvalue	18200
-------	
co2	0.7713
biomass	-0.5790
carbon dioxide emissions	0.2067
developing countries	0.1524

Eigenvalue	9625
-------	
developing countries	0.8235
biofuel	-0.4491
carbon dioxide emissions	-0.3355
biomass	0.0550

Term removal by generality, five iterations, most and least general term removed at each iteration:

Iteration 1 – Terms sorted by increasing level of generality

- *accelerated depreciation plan*
- *competitive electricity supply market*
- *decentralised development*
- *challenge 2008*
- *earth climate change*
- *delphi study*
- *baltic states*
- *commercial sector*
- *atmospheric carbon dioxide*
- *coal-fired power plants*
- *agricultural sector*
- *domestic renewable energy generators*
- *decarbonised world*
- *biogas system*
- *economic comparisons*
- *biomass*
- *developing countries*
- *co2*
- *biofuel*
- *carbon dioxide emissions*

Iteration 2 – Terms sorted by increasing level of generality

- *competitive electricity supply market*
- *decarbonised world*
- *domestic renewable energy generators*
- *decentralised development*
- *challenge 2008*
- *earth climate change*
- *delphi study*
- *baltic states*
- *commercial sector*
- *atmospheric carbon dioxide*
- *coal-fired power plants*
- *agricultural sector*
- *biogas system*
- *economic comparisons*
- *biofuel*
- *co2*
- *biomass*
- *developing countries*

Iteration 3 – Terms sorted by increasing level of generality

- *decarbonised world*
- *domestic renewable energy generators*
- *decentralised development*
- *challenge 2008*
- *earth climate change*
- *delphi study*
- *baltic states*
- *commercial sector*
- *coal-fired power plants*
- *atmospheric carbon dioxide*
- *agricultural sector*
- *biofuel*
- *biogas system*
- *economic comparisons*
- *co2*
- *biomass*

Iteration 4 – Terms sorted by increasing level of generality

- *domestic renewable energy generators*
- *decentralised development*
- *challenge 2008*
- *delphi study*
- *earth climate change*
- *baltic states*
- *coal-fired power plants*
- *commercial sector*
- *atmospheric carbon dioxide*
- *agricultural sector*
- *biofuel*
- *co2*
- *biogas system*
- *economic comparisons*

Iteration 5 – Terms sorted by increasing level of generality

- *decentralised development*
- *biogas system*
- *challenge 2008*
- *delphi study*
- *earth climate change*
- *baltic states*
- *coal-fired power plants*
- *commercial sector*
- *atmospheric carbon dioxide*
- *agricultural sector*
- *biofuel*
- *co2*

Final LSA concepts:

Eigenvalue	11065
-------	
biofuel	0.9975
agricultural sector	0.0415
coal-fired power plants	0.0378
commercial sector	0.0315

Eigenvalue	1582
-------	
agricultural sector	0.9898
commercial sector	0.1002
atmospheric carbon dioxide	0.0768
biofuel	-0.0456

Eigenvalue	3217
-------	
coal-fired power plants	0.9943
atmospheric carbon dioxide	0.0929
biofuel	-0.0420
commercial sector	0.0242

Eigenvalue	1201
-------	
atmospheric carbon dioxide	0.9902
coal-fired power plants	-0.0932
agricultural sector	-0.0861
commercial sector	0.0499

In the course of five iterations of data cleaning, the following terms were removed for being too general: *carbon dioxide emissions*, *developing countries*, *biomass*, *economic comparisons*, and *co2*. Similarly, these terms were removed for being too obscure: *accelerated depreciation plan*, *competitive electricity supply market*, *decarbonised world*, *domestic renewable energy generators*, and *decentralised development*. The results make sense. *Developing countries* and *economic comparisons*, for example, are general terms that could apply to any research field. *Carbon dioxide emissions* and *co2* are also hopelessly general topics. *Biomass* perhaps should not have been eliminated, but it does represent a

rather large subfield within renewable energy research. The terms eliminated due to obscurity also seem to be chosen appropriately. While they do belong in the realm of renewable energy, due to their specificity they do not often appear in documents containing other renewable energy-related terms. For example, Scirus finds only three documents that contain both terms *renewable energy* and *accelerated depreciation plan*; for comparison, it finds over one million containing *renewable energy* alone.

We now compare the LSA results from before and after term removal. Of all the highly weighted terms from the initial LSA concepts, only *biofuel* was not eliminated by data cleaning. This caused it to become a more significant term in the post-cleaning LSA, where it can be found as a highly weighted term in three concepts as opposed to the original two. It also seems that the concepts produced after data cleaning are more intuitive than those produced before cleaning. The first concept mostly represents the single term *biofuel*. Following this, the second concept contains *coal-fired power plants* and *atmospheric carbon dioxide*, which are clearly related. The fourth concept also weights these two terms the highest, although the order is reversed. The third concept contains *agricultural sector* and *commercial sector*, which are also clearly related. The concepts produced prior to term removal are more nonsensical, so it appears that this is an effective form of data cleaning.

4.5.2 Term removal by level of impact

The algorithm in the previous section used a heuristic to identify terms that are the most general and the most irrelevant to the topic, in order to remove those terms from the list. The algorithm we describe here uses a different metric and heuristic to identify terms for removal. Here, we identify and remove terms that have the largest and smallest impacts on the structure of the LSA concept vectors. The intuition behind removing terms with a large impact is that these are the terms which disrupt the output to a large extent and obscure the relationships between other terms in the set. In contrast, we may wish to remove terms with minimal impact on the concept vectors due to the fact that they are redundant in a sense and do not contribute unique information to the set of terms.

At the highest level, the algorithm proceeds as follows:

- Generate concept vectors from the complete set of terms.
- For each term, generate concept vectors from the set of all terms except for that particular term.
- For each term, compare this modified set of concept vectors with the original set of concept vectors, and use this comparison to assign an impact rating to the particular term that has temporarily been removed.
- Remove the terms with the highest and lowest impact ratings.
- Repeat the entire process as many times as desired.

The vast majority of the work this algorithm performs comes from the process of assigning a term an impact rating. This involves quantitatively comparing the concepts generated using all the terms in the set against the concepts generated without that term included. This is challenging for several reasons. First, if there are n terms in the complete set, each of the original concept vectors is n-

dimensional, whereas each of the modified concept vectors is $n-1$-dimensional. Second, we are comparing n original concept vectors with $n-1$ modified concept vectors. Third, the original concept vectors are not necessarily ordered in any meaningful way relative to the modified concept vectors. Properly comparing one set of vectors to the other thus requires first finding a (somewhat) optimal matching that associates each of the modified vectors with a particular vector from the original set. Finding this matching cannot require a prohibitively long runtime. And fourth, once we have a matching between the two sets of vectors, a similarity metric is needed to compare every matched pair of vectors and calculate the impact rating for the particular term being tested.

We overcome the first issue by simply ignoring the particular dimension of the original concept vectors represented by the term we are currently testing. For example, if the original vectors are of the form [*term 1*, *term 2*, *term 3*] and we are evaluating the impact of *term 2*, we re-represent the original vectors in the form [*term 1*, *term 3*]. The problem is now reduced to comparing n $n-1$-dimensional vectors against $n-1$ $n-1$-dimensional vectors.

We address the second and third issues simultaneously by seeking a matching of each of the $n-1$ modified vectors with $n-1$ of the n original vectors. Enumerating and evaluating all possibilities to find an optimal matching would require $O(n!)$ runtime, which is infeasible. Instead, a greedy algorithm is used that sacrifices optimality for polynomial runtime. The greedy algorithm iterates on the modified set of vectors, finding, for each one, an optimally matched vector from the set of original vectors, and then removing that original vector from possible use in subsequent iterations.

Determining which of the available original vectors is an optimal match for a given modified vector requires first addressing the fourth issue, namely, how we compare a single pair of vectors. This is a common mathematical question with several possible solutions. One metric for dissimilarity between two vectors is the angle between them; the larger the angle, the more dissimilar they are. Another possibility is to sum the absolute values of the differences between the components in one vector with the components in the other. A third option is to sum the squared differences between the components. Any of these methods could work for our purposes, so the choice is left as an adjustable parameter. Whichever method we choose, it is used to to quantitatively compare two vectors, allowing us to choose an optimally matched pair from a set of possibilities. Finally, once a matching between the $n-1$ modified vectors and $n-1$ of the n original vectors has been established, the vector dissimilarities are summed for each pair in the matching. This sum represents the level of impact that particular term has on the concept vectors.

Just as in the case with the algorithm described in the previous section for removing terms by level of generality, this heuristic is better applied if we perform several iterations, removing few terms each time. However, runtime is a more significant factor in this case than in the generality algorithm. Each iteration requires calculating a different set of eigenvectors corresponding to the removal of each term in the set and finding and evaluating a matching between that set of eigenvectors and the original set. For this reason, it is often the case that we choose to perform fewer iterations, removing more terms at each, in order to maintain a feasible runtime.

Finally, as in the case with the previously described algorithm, we can choose to remove terms that are highly impactful and/or minimally impactful at each iteration. We nearly always want to remove highly impactful terms, as they are the most disruptive to the structure of the LSA concepts. Depending on the situation, we may or may not choose to remove minimally impactful terms as well. These are redundant terms. If we wish to reduce the term list to only those that contribute unique meanings, we can remove them. If, on the other hand, we are interested in maintaining such terms, that can be done as well. The choice is left to the user as a parameter.

4.5.2.1 Software implementation and performance

To remove terms based on their level of impact, we first call *get_hit_counts* to extract all joint hit counts between terms. The function *remove_impactful_terms* is then called, which takes as its parameters a list of terms, the seed term, all hit counts, a number of iterations to execute, the number of impactful and non-impactful terms to remove at each iteration, and finally, an indication of which mode of vector comparison to use.

For each iteration, the function first produces a covariance matrix from the hit counts and calculates their eigenvectors. Then, for each term, the function removes that term from the matrix, recalculates the eigenvectors, finds a matching between the original eigenvectors and the modified eigenvectors, and sums the vector difference between all vectors in the matching, using the vector comparison mode indicated by the user. The vector comparison modes (angle, sum of absolute difference, sum of squared difference) are implemented in a modular fashion as separate functions that take two vectors as input and produce a distance as output. In this way, other distance metrics can easily be added as well.

Again, as with the functions *lsa* and *remove_general_terms*, the performance bottleneck is hit count extraction. However, the runtime of the analysis portion plays a more significant role in this algorithm than the others. For each iteration, each term is tested individually. For each term, a set of eigenvectors is calculated, which costs $O(n^3)$ time. A matching between these eigenvectors and the originals is then found and evaluated. This also rquires $O(n^3)$ time, since for each of the $O(n^2)$ pairs of vectors between the two sets, each of the $O(n)$ values is read. The overall runtime of the analysis portion of this algorithm (the portion after hit counts have already been extracted) is therefore $O(i \times n^4)$, for i iterations.

After all hit counts have been extracted, running 10 single-term-removal iterations on a list of 15 terms takes approximately 1 second. With 50 terms, this takes approximately 80 seconds.

4.5.2.2 Results

This section presents results for term removal based on level of impact. We use the same terms as those which are used in the demonstration of term removal by generality in section 4.5.1.2. We first present LSA concepts of the original list. (Note that these original concepts are identical to those shown in section 4.5.1.2.) Five iterations of term removal are then performed and displayed, deleting the most

and least impactful term from the list at each iteration, using the absolute difference metric for vector distance. Afterwards, LSA is again performed on the cleaned list of terms. All LSA results are abridged, displaying the four highest-weighted terms of the four highest-weighted concepts.

Initial LSA concepts:

Eigenvalue	29529
-------	
biomass	0.7743
co2	0.5115
biofuel	0.2822
developing countries	0.1856

Eigenvalue	10746
-------	
biofuel	0.5899
developing countries	0.5060
carbon dioxide emissions	0.4783
co2	-0.3561

Eigenvalue	18200
-------	
co2	0.7713
biomass	-0.5790
carbon dioxide emissions	0.2067
developing countries	0.1524

Eigenvalue	9625
-------	
developing countries	0.8235
biofuel	-0.4491
carbon dioxide emissions	-0.3355
biomass	0.0550

Term removal by impact, five iterations, most and least impactful term removed at each iteration:

Iteration 1 – Terms sorted by increasing level of impact

- *biofuel*
- *carbon dioxide emissions*
- *co2*
- *developing countries*
- *biomass*
- *economic comparisons*
- *biogas system*
- *agricultural sector*
- *domestic renewable energy generators*
- *decarbonised world*
- *commercial sector*
- *atmospheric carbon dioxide*
- *coal-fired power plants*
- *baltic states*
- *earth climate change*
- *challenge 2008*
- *competitive electricity supply market*
- *delphi study*
- *accelerated depreciation plan*
- *decentralised development*

Iteration 2 – Terms sorted by increasing level of impact

- *co2*
- *developing countries*
- *biomass*
- *carbon dioxide emissions*
- *biogas system*
- *economic comparisons*
- *atmospheric carbon dioxide*
- *commercial sector*
- *agricultural sector*
- *baltic states*
- *coal-fired power plants*
- *domestic renewable energy generators*
- *decarbonised world*
- *delphi study*
- *challenge 2008*
- *competitive electricity supply market*
- *accelerated depreciation plan*
- *earth climate change*

Iteration 3 – Terms sorted by increasing level of impact

- *developing countries*
- *carbon dioxide emissions*
- *economic comparisons*
- *biogas system*
- *domestic renewable energy generators*
- *decarbonised world*
- *biomass*
- *commercial sector*
- *atmospheric carbon dioxide*
- *delphi study*
- *coal-fired power plants*
- *agricultural sector*
- *baltic states*
- *accelerated depreciation plan*
- *competitive electricity supply market*
- *challenge 2008*

Iteration 4 – Terms sorted by increasing level of impact

- *economic comparisons*
- *biogas system*
- *agricultural sector*
- *coal-fired power plants*
- *carbon dioxide emissions*
- *biomass*
- *atmospheric carbon dioxide*
- *commercial sector*
- *decarbonised world*
- *domestic renewable energy generators*
- *delphi study*
- *baltic states*
- *accelerated depreciation plan*
- *competitive electricity supply market*

Iteration 5 – Terms sorted by increasing level of impact

- *agricultural sector*
- *atmospheric carbon dioxide*
- *carbon dioxide emissions*
- *commercial sector*
- *biomass*
- *delphi study*
- *baltic states*
- *domestic renewable energy generators*
- *decarbonised world*
- *accelerated depreciation plan*
- *coal-fired power plants*
- *biogas system*

Final LSA concepts:

Eigenvalue	24517
-------	
biomass	0.9975
carbon dioxide emissions	0.0572
coal-fired power plants	0.0290
atmospheric carbon dioxide	0.0224

Eigenvalue	3151
-------	
coal-fired power plants	0.9920
carbon dioxide emissions	-0.0949
atmospheric carbon dioxide	0.0762
biomass	-0.0255

Eigenvalue	9699
-------	
carbon dioxide emissions	0.9931
coal-fired power plants	0.0899
biomass	-0.0608
atmospheric carbon dioxide	0.0430

Eigenvalue	1190
-------	
atmospheric carbon dioxide	0.9939
coal-fired power plants	-0.0816
commercial sector	0.0605
carbon dioxide emissions	-0.0375

The following terms were removed for their low level of impact: *biofuel*, *co2*, *developing countries*, *economic comparisons*, and *agricultural sector*. These are similar to the high-generality terms identified in the previous section. C*O2*, *developing countries*, and *economic comparison* appear in both

lists, and *biofuel* appears here in place of the similar term *biomass* from the high-generality list. It seems that the inclusion of general terms tends to have a small impact on the resultant LSA concepts.

These terms were removed for their high level of impact: *decentralised development*, *earth climate change*, *challenge 2008*, *competitive electricity supply market*, and *biogas system*. These terms bear some resemblance to the low-generality terms identified previously. Specifically, *decentralised development* and *competitive electricity supply market* are identified using both metrics.

After removing high- and low- impact terms, we observe strikingly clean LSA concepts. The first concept mostly represents *biomass*, with some contribution from the "pollution group" of *carbon dioxide emissions*, *coal-fired power plants*, and *atmospheric carbon dioxide*. The second and third concepts consist of these same four terms, with a decreasing contribution from *biomass* in each successive concept. Finally, in the fourth concept, *biomass* is no longer in the top four of term weightings, and is replaced by *commercial sector*, but *carbon dioxide emissions*, *coal-fired power plants*, and *atmospheric carbon dioxide* remain prevalent in all four concepts. Term removal by level of impact appears to be at least as effective as term removal by generality as a method of data cleaning.

4.6 Latent Semantic Analysis wrap-up

This chapter has discussed the statistical technique known as Latent Semantic Analysis and how we apply it to reveal the concepts that underlie a set of terms. We described the well-known algorithm as well as our novel enhancement that allows us to take advantage of the immense volume of publications indexed by the search engines we utilize. We have also discussed two data cleaning algorithms that we designed to identify and remove terms that reduce the quality of the LSA concept vectors.

In the next chapter, we demonstrate how all of our software modules, including keyword extraction, hit count extraction, growth rate analysis, and LSA, can be used in conjunction to analyze a field of research.

Chapter 5: Renewable Energy Case Study Results

5.1 Overview

Chapters 3 and 4 presented methods and results for each piece of our methodology. We demonstrated adjusting the available parameters of each operation, and we arrived at conclusions as to which parameter values best suit our needs. In this chapter, a single, real-world example is presented that utilizes every piece of our methodology. For every operation, we use what we believe is the best parameter setting. (In some cases, the best choice of parameters may involve merging results from different settings.) As per our case study, the example presented in this chapter analyzes the field of renewable energy.

First, related keywords are collected. Second, hit counts are extracted for each term over the range of years from 2005 through 2008. Third, the hit counts are used to rank the terms by growth rate and to condense the list to only fast-growing terms. Fourth, LSA is performed on the reduced set of terms. And finally, fifth, growth rates are calculated again, but in this case for entire concepts rather than for individual terms. This process is illustrated in the diagram below.

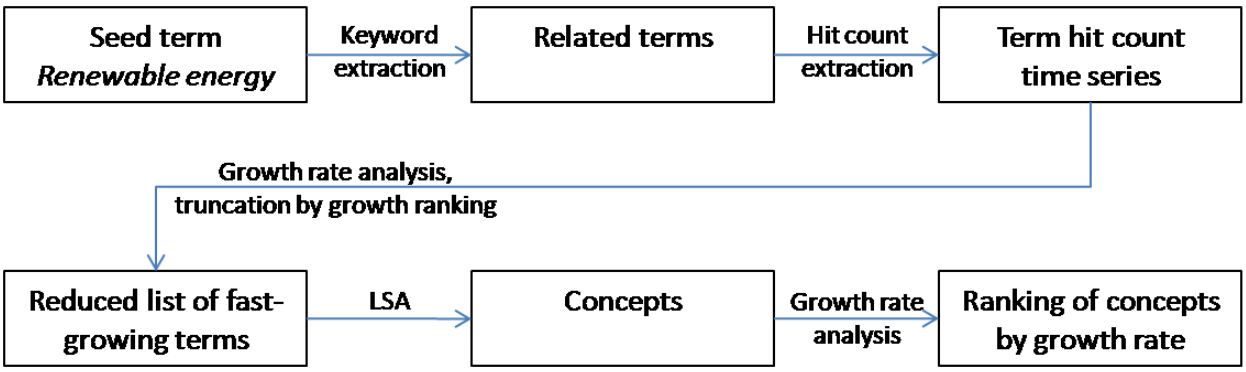

Figure 13. Process combining all modules for analysis of the renewable energy field.

5.2 Keyword extraction

To acquire a diverse set of terms, we combine several methods of keyword extraction. Terms from the first fifty articles of both Engineering Village databases, Compendex and Inspec, are collected. Only the uncontrolled keywords from these databases are considered, as controlled terms tend to be overly general. Next, terms are collected from Scirus. The hierarchy extension pattern of [2, 2, 2] is used in this operation, as we find that this strikes a good balance between relevance and specificity. Finally, we also use back-pointing term generation. 5,000 candidate terms are tested for back-pointing to *renewable energy*. Only terms that point back through one level are considered, because with such a large number of candidate terms, the set of two-level terms contains a large proportion that are irrelevant. However, we find that the number of one-level terms is too low. We therefore increase the number of back-pointing terms by also including all terms that point back through one level to the synonymous term *sustainable energy*.

The uncontrolled terms from Compendex and Inspec are listed in the appendix in section A.1. The terms collected from Scirus using the [2, 2, 2] pattern are presented in section A.2. And the back-pointing terms, for both *renewable energy* and *sustainable energy*, are in section A.3. After combining the term sets and removing duplicates from Compendex, Inspec, Scirus, back-pointing to *renewable energy*, and back-pointing to *sustainable energy*, we are left with a list of 1,343 terms. The merged term list is presented in its entirety in appendix A.4. A small sample of some of its high-quality terms is displayed here.

Sample of renewable energy-related terms:

- *amorphous silicon solar cells*
- *biogas electricity conversion*
- *biomass gasification*
- *clean coal technologies*
- *district heating*
- *dual fuel engine*
- *electrodialysis*
- *energy crops*
- *flexible photo-voltaic solar cell*
- *fusion energy*
- *geothermal heat pumps*
- *heat exchangers*
- *hybrid development*
- *hybrid solar-wind system*
- *hydrogen fuel*
- *microhydro*
- *plasma physics*
- *solar heat pump combination*
- *solar hot water system*
- *stingray tidal stream generator*
- *thermochemical conversion*
- *tidal power*
- *wave energy conversion*
- *wind energy conversion systems*

This list contains many good, promising terms. *Amorphous silicon solar cells*, *biomass gasification*, *dual fuel engine*, *flexible photo-voltaic solar cell*, *geothermal heat pumps*, *hybrid solar-wind system*, *microhydro*, *solar heat pump combination*, and *stingray tidal stream generator*, for example, represent specific renewable energy technologies.

On the other hand, the list contains some irrelevant terms such as *1.27e15 j*, *450*, *[jel] q48*, *bachelor degree program*, *economic topic*, and *top-down*. It also contains some overly general terms such as *energy*, *environment issues*, *heating*, and *renewable sources*. (These terms are not included in the abridged list above, but they appear in the appendix.)

5.3 Hit count extraction

Hit counts are extracted for all terms over the range of years from 2005 through 2008. This relatively short range was chosen because we are most interested in technologies that are just beginning to become known. A ten-year range, for example, would automatically exclude terms with no hits prior to 2005, which is contrary to our goal. It should also be noted that we do not include the current year, 2009, because hit counts for the current year would be artificially low.

These hit counts are extracted with *renewable energy* included in the query to assure that each term is counted in the appropriate context. We use Scirus, rather than Compendex, for the larger hit counts it produces.

For space considerations, hit counts are displayed for only the first twenty terms.

Hit count extraction: Scirus, 2005-2008, *renewable energy* in query (abridged to twenty terms):

	2005	2006	2007	2008
1.27e15 j	0	0	0	0
450	4776	7906	22045	9155
[jel] q48	2	6	5	4
aboriginal communities	179	340	754	307
accelerated depreciation plan	1	1	1	0
accelerator	583	1057	15413	1914
access key	27	30	983	819
active power	202	311	901	222
actuators	398	907	2829	2035
ad 2020-2021	0	0	0	0
advisory committee	2735	4301	12461	4347
aerodynamic	579	960	3623	1196
aerodynamics	492	837	4845	1471
after sales service	70	100	905	151
aggregated wind turbines	0	2	1	1
agricultural sector	1045	1741	4227	1600
air infiltration	176	289	500	150
alarm limit	2	1	7	7
alloy surfaces	6	13	48	15
alternative energy	3269	5788	50890	21502

A relatively wide range of hit counts is observed. The inclusion of *renewable energy* in the search query has clearly helped reduce the hit counts of irrelevant terms. *Alarm limit*, for example, would have had 282 hits in 2005 if not for the inclusion of *renewable energy*, which brought the hit count down to 2. On the other hand, the inclusion of *renewable energy* only decreased the hit count of the more relevant *alternative energy* by approximately a factor of three, as it would have had 9,483 hits otherwise.

5.4 Growth rate analysis of terms

For growth rate analysis, we consider two different metrics, namely the ratio of the final and initial hit count, and the exponential curve-fitting parameter. These metrics most closely meet our interest of identifying terms with large early-stage growth. We also normalize all hit counts to *renewable energy*'s hit counts, to account for year-to-year fluctuations in the total number of articles Scirus indexes.

Appendix A.6 presents the complete ranked list of growth rates for both metrics. Abridged lists are displayed here, with the fifty highest- and lowest-ranked terms from each metric.

Normalized ratio, top 50 terms:		Normalized exponential rate, top 50 terms:	
space technologies	14.595	*renewable energy policy harmonisation*	4.271
dinosaurs	13.569	*optimal renewable energy mathematical model*	4.161
challenge 2008	7.643	*ecologically optimized transport systems*	4.051
book reviews	5.207	*energy-saving technical progress*	3.167
renewable energy incentives	4.638	*tourist ventures*	3.131
access key	4.606	*centre of appropriate technology*	3.084
Nanotechnology	4.066	*load powering*	2.992
Microhydro	3.978	*renewable energy promotion strategies*	2.992
pv-compact	3.948	*accelerated depreciation plan*	2.923
Donate	3.706	*renewable energy development strategies*	2.813
petroleum resources	3.589	*renewable energy generation cost*	2.784
Environmentalism	3.402	*optimal renewable energy mathematical (orem) model*	2.784
european islands	3.189	*emission control policy*	2.754
editorials	3.156	*household energy situation*	2.715
testing phase	3.022	*carbon dioxide discharges*	2.664
chemical safety	2.87	*hybrid diesel systems*	1.944
earth science	2.677	*decentralized power systems penetration*	1.783
vibration control	2.405	*intermittent renewable energy penetration*	1.783
chronic beryllium disease	2.342	*decarbonised world*	1.575
beryllium disease	2.255	*biogas electricity conversion*	1.14
food technology	2.205	*domestic technology manufacturing*	1.14
energy politics	2.182	*renewable energy engineering curriculum*	1.14
articles	2.114	*renewable energy engineering system*	1.14
alternative fuels	1.973	*workshop in a box program*	1.14
earth climate change	1.949	*dinosaurs*	1.14
global warming	1.93	*commercial energy scene*	1.071
wind systems	1.892	*indian national programme*	1.071
energy technology	1.886	*market adoptions*	1.071
biofuel	1.744	*space technologies*	1.033
environmental engineering	1.664	*book reviews*	0.874
greenhouse effect	1.611	*nanotechnology*	0.609
carbon cap-and-trade	1.603	*access key*	0.599
wind energy	1.573	*challenge 2008*	0.592
carbon emissions	1.568	*pv-compact*	0.501
biofuels	1.567	*microhydro*	0.495
failure analysis	1.546	*petroleum resources*	0.478
biofuel energy	1.495	*donate*	0.457
renewable energy research	1.411	*vibration control*	0.439
research output	1.408	*chronic beryllium disease*	0.424
solar water	1.394	*editorials*	0.423

marine renewable energy	1.39	***beryllium disease***	0.402
offshore wind turbine	1.329	***renewable energy incentives***	0.368
ignalina nuclear power plant	1.306	***alternative fuels***	0.363
Supercomputing	1.291	***chemical safety***	0.356
Volcano	1.289	***earth science***	0.349
Estonia	1.237	***articles***	0.316
continuous energy supply	1.215	***global warming***	0.315
gibraltar strait	1.215	***environmentalism***	0.314
legal environment	1.195	***wind systems***	0.299
Publications	1.181	***energy technology***	0.297

Normalized ratio, bottom 50 terms:		**Normalized exponential rate, bottom 50 terms:**	
intrinsic technology problem	0	***uk nffo***	***-0.988***
kirklees councils	0	***thermal waters***	***-1.01***
load powering	0	***pressure decline***	***-1.036***
local employments	0	***rural energy enterprises***	***-1.039***
low activity subjects	0	***peak energies***	***-1.045***
marginalized mountain population	0	***broad policy guidelines***	***-1.066***
market adoptions	0	***solar regime***	***-1.104***
multilevel search engines	0	***lighting end-uses***	***-1.12***
natural gas supplies reduction	0	***optimal renewable energy model***	***-1.132***
nonpooled electricity trading	0	***overview of fuel cell***	***-1.162***
optimal renewable energy mathematical (orem) model	0	***subsidy measures***	***-1.18***
optimal renewable energy mathematical model	0	***elsevier (co)***	***-1.201***
petroleum crude oil consumption	0	***renewable energy systems cost***	***-1.204***
policies and measures to enhance use of renewable energy sources	0	***solar heat pump combination***	***-1.209***
power plant financing process	0	***web-based renewable energy courses***	***-1.238***
project datums	0	***cascade of hydro power plants***	***-1.279***
public lands -- oregon -- prineville region – management	0	***future energy supply networks***	***-1.279***
recursive linear planning model	0	***local employments***	***-1.348***
recursive linear program	0	***policies and measures to enhance use of renewable energy sources***	***-1.348***
regional power market competition	0	***solar pv utilization***	***-1.348***
regulated rural energy concessions	0	***policies formulation***	***-1.369***
renewable energy development strategies	0	***china-india cooperation***	***-1.423***
renewable energy engineering curriculum	0	***rural electrification in senegal***	***-1.734***
renewable energy engineering system	0	***energy demand variations***	***-1.844***

renewable energy generating equipments	0	***stingray tidal current generator***	***-1.913***
renewable energy generation cost	0	***german renewable energy source act***	***-2.052***
renewable energy policy harmonisation	0	***renewable energy distribution***	***-2.404***
renewable energy product purchase	0	***renewable energy company limited***	***-2.766***
renewable energy promotion strategies	0	***ministry of economic affairs of taiwan***	***-3.366***
renewable energy systems reliability	0	***underwater power generation***	***-3.488***
renewable energy trading experience	0	***greenhouse climate control***	***-3.535***
renewable facilities	0	***liquid dampers***	***-3.655***
ring-fenced markets	0	***bureau of energy of ministry of economical affairs***	***-3.696***
social acceptance variation	0	***conventional generation cost***	***-3.696***
socio economic optimal renewable energy model	0	***national power installation capacity***	***-3.696***
solar heat pump combination	0	***policy design impacts***	***-3.696***
solar pv utilization	0	***postgraduate university courses***	***-3.696***
sustained renewable energy deployment	0	***renewable energy equipment procurement***	***-3.696***
tax revenue method	0	***renewable energy systems allocation***	***-3.696***
tertiary renewable energy education	0	***dynamic life cycle assessment***	***-3.758***
tertiary trained engineers	0	***rd and d***	***-3.765***
tertiary trained policy makers	0	***education and trainings***	***-3.835***
tertiary trained scientists	0	***emissions (asphalt)***	***-3.904***
time 2020 year to 2021 year	0	***negative secondary impacts***	***-3.904***
tourist ventures	0	***renewable energy plan 2012***	***-3.904***
uniform green power market mode	0	***undergraduate university courses***	***-3.944***
web-based renewable energy courses	0	***aggregated wind turbines***	***-3.973***
wind excitation	0	***regulatory law framework***	***-4.767***
windfarm market	0	***eirev***	***-4.975***
workshop in a box program	0	***tlcd***	***-4.975***

The lists of high-growth terms include some that are promising, for example *ecologically optimized transport systems*, *microhydro*, *nanotechnology*, *biofuel*, and *biogas electricity conversion*. However, we also see some puzzling terms, such as *dinosaurs* and *book reviews*. For an unknown reason, these terms had particularly large hit counts from Scirus in 2008, even when paired with *renewable energy*. This suggests that hit counts do not always function perfectly as an indication of technological prevalence.

In the group of low-growth terms, we see very positive results. Most of the terms in this group belong there, due to being overly general and uninteresting, including *intrinsic technology problem*, *power plant financing process*, and *renewable energy distribution*.

5.5 Latent Semantic Analysis

Because LSA requires extracting joint hit counts for every pair of terms, it is only feasible with a relatively small number of terms. Furthermore, even if runtime were not an issue, LSA concepts become increasingly difficult to interpret as the number of terms increases. For this reason, the growth rate analysis from the previous step is used to significantly reduce the size of the term list. LSA is then run on this small set of fast-growing terms to identify the concepts that underlie them. We choose to include the fifty highest-rated terms from each of our two growth metrics. These terms are listed in the previous section but are reproduced here after the two sets of terms have been merged by removing duplicates and sorting the remaining terms alphabetically. After merging, 77 terms remain, to be input into the LSA algorithm.

Highest-growth terms, for use in LSA:

- *accelerated depreciation plan*
- *access key*
- *alternative fuels*
- *articles*
- *beryllium disease*
- *biofuel*
- *biofuel energy*
- *biofuels*
- *biogas electricity conversion*
- *book reviews*
- *carbon cap-and-trade*
- *carbon dioxide discharges*
- *carbon emissions*
- *centre of appropriate technology*
- *challenge 2008*
- *chemical safety*
- *chronic beryllium disease*
- *commercial energy scene*
- *continuous energy supply*
- *decarbonised world*
- *decentralized power systems penetration*
- *dinosaurs*
- *domestic technology manufacturing*
- *donate*
- *earth climate change*
- *earth science*
- *ecologically optimized transport systems*
- *editorials*
- *emission control policy*
- *energy politics*
- *energy technology*
- *energy-saving technical progress*
- *environmental engineering*
- *environmentalism*
- *estonia*
- *european islands*
- *failure analysis*
- *food technology*
- *gibraltar strait*
- *global warming*
- *greenhouse effect*
- *household energy situation*
- *hybrid diesel systems*
- *ignalina nuclear power plant*
- *indian national programme*
- *intermittent renewable energy penetration*
- *legal environment*
- *load powering*
- *marine renewable energy*
- *market adoptions*
- *microhydro*
- *nanotechnology*
- *offshore wind turbine*
- *optimal renewable energy mathematical (orem) model*
- *optimal renewable energy mathematical model*
- *petroleum resources*
- *publications*
- *pv-compact*
- *renewable energy development strategies*
- *renewable energy engineering curriculum*
- *renewable energy engineering system*
- *renewable energy generation cost*
- *renewable energy incentives*
- *renewable energy policy harmonisation*
- *renewable energy promotion strategies*
- *renewable energy research*
- *research output*
- *solar water*
- *space technologies*
- *supercomputing*
- *testing phase*
- *tourist ventures*
- *vibration control*
- *volcano*
- *wind energy*
- *wind systems*
- *workshop in a box program*

Although 77 terms is not prohibitively large, LSA concepts on such a set would be difficult to interpret. We have seen that our data cleaning methods are effective at removing terms that are

disruptive to the results, so we use these methods here. Term removal by level of impact is applied first, eliminating the most impactful term on each of ten iterations. This step only serves to remove disruptive terms. We allow low-impact terms to remain for now, leaving them for the generality-based cleaning to be performed later. The results of impact-based cleaning are shown here.

Removal of most impactful term, ten iterations:

Terms removed at each iteration:

- *tourist ventures*
- *volcano*
- *continuous energy supply*
- *food technology*
- *accelerated depreciation plan*
- *marine renewable energy*
- *publications*
- *gibraltar strait*
- *supercomputing*
- *solar water*

Terms remaining after ten iterations:

- *book reviews*
- *earth science*
- *optimal renewable energy mathematical (orem) model*
- *workshop in a box program*
- *space technologies*
- *research output*
- *renewable energy incentives*
- *indian national programme*
- *access key*
- *testing phase*
- *chemical safety*
- *emission control policy*
- *renewable energy development strategies*
- *ecologically optimized transport systems*
- *centre of appropriate technology*
- *carbon dioxide discharges*
- *load powering*
- *renewable energy research*
- *microhydro*
- *renewable energy promotion strategies*
- *petroleum resources*
- *renewable energy engineering curriculum*
- *chronic beryllium disease*
- *biogas electricity conversion*
- *renewable energy policy harmonisation*
- *intermittent renewable energy penetration*
- *pv-compact*
- *renewable energy engineering system*
- *decentralized power systems penetration*
- *energy-saving technical progress*
- *failure analysis*
- *decarbonised world*
- *domestic technology manufacturing*
- *carbon cap-and-trade*
- *editorials*
- *beryllium disease*
- *energy politics*
- *biofuel energy*
- *household energy situation*
- *vibration control*
- *offshore wind turbine*
- *commercial energy scene*
- *optimal renewable energy mathematical model*
- *ignalina nuclear power plant*
- *renewable energy generation cost*
- *market adoptions*
- *dinosaurs*
- *hybrid diesel systems*
- *challenge 2008*
- *wind systems*
- *legal environment*
- *european islands*
- *earth climate change*
- *environmentalism*
- *estonia*
- *environmental engineering*
- *wind energy*
- *greenhouse effect*
- *carbon emissions*
- *nanotechnology*
- *alternative fuels*
- *global warming*
- *biofuel*
- *energy technology*
- *biofuels*
- *articles*
- *donate*

Of the ten terms that were removed, three are surprising: *volcano* (which is relevant to geothermal energy), *marine renewable energy*, and *solar water*. These terms seem closely tied to others in the list, so it is unclear why they are highly impactful. The rest of the terms that were removed, however, seem appropriate.

Following impact-based data cleaning, we perform another round of cleaning based on level of generality. We are interested in eliminating terms that are either too general or too specific, so five iterations are executed, removing the most and least general term at each iteration. Results are shown here.

Removal of most and least general term, five iterations:

High generality terms removed at each iteration:

- *nanotechnology*
- *alternative fuels*
- *greenhouse effect*
- *dinosaurs*
- *renewable energy engineering curriculum*

Low generality terms removed at each iteration:

- *european islands*
- *ecologically optimized transport systems*
- *renewable energy policy harmonisation*
- *carbon dioxide discharges*
- *challenge 2008*

Terms remaining after five iterations:

- *ignalina nuclear power plant*
- *energy-saving technical progress*
- *earth climate change*
- *carbon cap-and-trade*
- *biofuel energy*
- *offshore wind turbine*
- *energy politics*
- *renewable energy generation cost*
- *hybrid diesel systems*
- *access key*
- *research output*
- *vibration control*
- *centre of appropriate technology*
- *estonia*
- *testing phase*
- *pv-compact*
- *chronic beryllium disease*
- *beryllium disease*
- *renewable energy incentives*
- *microhydro*
- *renewable energy development strategies*
- *wind systems*
- *donate*
- *articles*
- *earth science*
- *load powering*
- *legal environment*
- *wind energy*
- *space technologies*
- *market adoptions*
- *global warming*
- *biogas electricity conversion*
- *petroleum resources*
- *failure analysis*
- *editorials*
- *energy technology*
- *book reviews*
- *carbon emissions*
- *biofuels*
- *biofuel*
- *chemical safety*
- *indian national programme*
- *optimal renewable energy mathematical model*
- *workshop in a box program*
- *environmentalism*
- *emission control policy*

- *renewable energy research*
- *decarbonised world*
- *environmental engineering*
- *decentralized power systems penetration*
- *household energy situation*
- *optimal renewable energy mathematical (orem) model*
- *commercial energy scene*
- *renewable energy promotion strategies*
- *intermittent renewable energy penetration*
- *renewable energy engineering system*
- *domestic technology manufacturing*

With the exception of *nanotechnology*, the terms labeled as highly general appear to have been properly identified. (It is difficult, however, to say whether *dinosaurs* is general or not to the field of renewable energy, since for the most part, it is irrelevant.) The low generality terms also appear to agree with intuition. It is not surprising, for example, that *ecologically optimized transport systems* is too specific to be a large component of many concepts.

After having removed several terms that are highly impactful, general, or specific, we then perform LSA on the remaining 57 terms. LSA produces 57 concepts, each consisting of weighted combinations of all 57 terms. In the results of this particular experiment, after the 37th concept, all eigenvalues are essentially zero, meaning those concepts are not representative of the data. Therefore, for space considerations, these concepts are not displayed. We also conserve space by only displaying the highly-weighted terms in each concept. In appendix section A.6, the first 37 concepts are displayed, each with the ten highest-weighted terms. An abridged version is displayed here, containing the first ten concepts, with the five highest-weighted terms of each.

LSA concepts:

Eigenvalue	86559	**Eigenvalue**	29711
-------		***-------***	
global warming	0.59149	***carbon emissions***	0.49354
articles	0.53163	***articles***	-0.46009
energy technology	0.41600	***global warming***	0.39632
wind energy	0.37611	***biofuels***	0.34675
carbon emissions	0.14663	***wind energy***	-0.33847

Eigenvalue	39996	**Eigenvalue**	25348
-------		***-------***	
wind energy	0.68584	***energy technology***	-0.64540
articles	-0.56990	***biofuels***	0.40326
global warming	0.20395	***articles***	0.37028
environmental engineering	-0.18486	***global warming***	0.22892
carbon emissions	-0.16017	***biofuel***	0.20243

Eigenvalue	23308
-------	
biofuels	-0.67133
global warming	0.43221
donate	0.29221
biofuel	-0.26514
energy technology	-0.26326

Eigenvalue	9365
-------	
biofuel	-0.72881
carbon emissions	-0.31870
biofuels	0.31370
book reviews	0.27965
environmental engineering	0.27859

Eigenvalue	15788
-------	
carbon emissions	-0.50139
wind energy	-0.39979
energy technology	0.36495
environmental engineering	-0.33807
global warming	0.32736

Eigenvalue	9115
-------	
environmental engineering	0.62995
carbon emissions	-0.53592
biofuel	0.42983
environmentalism	0.22101
petroleum resources	-0.11971

Eigenvalue	13033
-------	
donate	-0.79396
earth science	-0.39650
global warming	0.28625
energy technology	-0.19918
biofuels	-0.15382

Eigenvalue	6149
-------	
earth science	-0.50230
petroleum resources	-0.45964
space technologies	-0.43887
donate	0.34262
energy technology	0.25692

Unfortunately, these results are quite difficult to interpret. *Articles*, *donate*, and *book reviews* are noise terms which were not identified by the two stages of data cleaning. But even when ignoring these terms, it is difficult to see strong connections among the terms within each concept. We have found that even though mathematically speaking, these concepts are the principal components of the articles indexed by Scirus, it can sometimes be difficult to acquire useful qualitative information from LSA results.

5.6 Growth rate analysis of concepts

LSA produces results that are valuable in their own right, where its concept vectors can illuminate important interrelationships between terms. However, LSA can also be utilized in a different way. We are interested in technologies with large growth rates. But as we have discussed, sometimes simply querying the term itself does not give an accurate hit count, due to synonymy, etc. LSA provides a solution to this problem by allowing us to calculate and rank the growth rates of entire concepts.

This section displays results for concept growth ranking using the 37 concepts of ten terms each that were calculated in the previous section and that appear in appendix A.6. We first display the yearly hit counts for each concept. A concept's hit count is calculated by summing the appropriately weighted individual hit counts of the terms it encompasses. For example, if we have a concept which is comprised

of 0.8 *wind* and 0.6 *solar*, and if the hit counts for *wind* and *solar* in 2005 are 100 and 200, respectively, then the hit count for the entire concept in 2005 is $0.8 \times 100 + 0.6 \times 200 = 200$. If a term has a negative weighting in a concept, its absolute value is considered, so that its hit counts contribute positively to the concept as a whole.

Hit counts of LSA concepts:

	2005	2006	2007	2008
Concept 1	9163	17340	401338	114805
Concept 2	8333	15725	358056	101873
Concept 3	8177	15527	307870	99216
Concept 4	6420	13000	258913	83044
Concept 5	6550	12731	278383	79779
Concept 6	6614	12276	257893	77310
Concept 7	3864	7538	160537	55042
Concept 8	2673	5465	86109	31116
Concept 9	2581	5014	43501	29788
Concept 10	1953	4164	97483	30684
Concept 11	2085	4191	102991	30887
Concept 12	1907	3919	114786	23337
Concept 13	1587	3192	85872	18066
Concept 14	1488	2727	69847	18041
Concept 15	569	1168	22295	9457
Concept 16	574	1180	14536	7454
Concept 17	885	1714	21837	12810
Concept 18	1636	3025	40188	21320
Concept 19	422	881	17047	5462

	2005	2006	2007	2008
Concept 20	512	982	75478	8300
Concept 21	110	218	3174	1574
Concept 22	181	372	3503	3376
Concept 23	295	637	6892	3702
Concept 24	295	616	11394	3830
Concept 25	134	290	2096	1983
Concept 26	186	379	3804	2262
Concept 27	134	237	4631	1398
Concept 28	246	482	3022	2689
Concept 29	123	254	1996	3119
Concept 30	65	135	1108	1007
Concept 31	70	167	2154	1081
Concept 32	198	474	14034	6670
Concept 33	96	193	1681	1556
Concept 34	33	70	1054	314
Concept 35	17	39	820	235
Concept 36	28	41	1659	424
Concept 37	1	2	24	0

These hit counts are now used to calculate a growth rate for each concept. We use the ratio metric, while normalizing hit counts against *renewable energy*. The results of this are shown here.

Growth rate of concepts, normalized ratio:

Concept 32	5.116	***Concept 36***	2.289	***Concept 24***	1.970
Concept 29	3.861	***Concept 11***	2.249	***Concept 4***	1.964
Concept 22	2.835	***Concept 25***	2.249	***Concept 19***	1.963
Concept 15	2.523	***Concept 17***	2.198	***Concept 23***	1.908
Concept 33	2.467	***Concept 21***	2.172	***Concept 1***	1.903
Concept 20	2.460	***Concept 7***	2.163	***Concept 12***	1.859
Concept 10	2.385	***Concept 35***	2.141	***Concept 2***	1.856
Concept 30	2.366	***Concept 18***	1.979	***Concept 26***	1.851
Concept 31	2.349	***Concept 16***	1.973	***Concept 5***	1.850

Concept 3	1.842	***Concept 9***	1.753	***Concept 34***	1.436
Concept 14	1.841	***Concept 13***	1.729	***Concept 37***	0.000
Concept 6	1.775	***Concept 28***	1.661		
Concept 8	1.768	***Concept 27***	1.589		

The concepts are listed in the appendix section A.6, but the five highest- and lowest-ranked concepts are reproduced here for convenience.

Five fastest-growing concepts:

Concept 32	
Eigenvalue	235
-------	
space technologies	0.71929
petroleum resources	-0.68795
biofuel energy	0.07424
research output	0.03859
access key	-0.03380
carbon cap-and-trade	0.02383
failure analysis	-0.01755
legal environment	0.01275
energy technology	-0.00617
renewable energy incentives	-0.00593

Concept 29	
Eigenvalue	359
-------	
renewable energy incentives	-0.70381
microhydro	0.69937
carbon cap-and-trade	-0.10363
legal environment	-0.05095
failure analysis	0.02835
beryllium disease	-0.01741
chronic beryllium disease	-0.01721
pv-compact	-0.01634
chemical safety	0.01403
offshore wind turbine	0.01397

Concept 22	
Eigenvalue	1164
-------	
testing phase	0.97266
environmentalism	-0.17111
energy politics	-0.14771
chemical safety	-0.03003
failure analysis	0.02701
book reviews	-0.01476
Biofuels	0.01388
legal environment	0.01286
beryllium disease	-0.01212
chronic beryllium disease	-0.01205

Concept 15	
Eigenvalue	3630
-------	
microhydro	0.64025
renewable energy incentives	0.63675
chemical safety	-0.24479
editorials	-0.17378
failure analysis	-0.17070
legal environment	-0.16339
renewable energy research	0.11510
environmentalism	0.06372
articles	-0.05679
book reviews	-0.05040

Concept 33	
Eigenvalue	166
-------	
access key	0.72416
research output	-0.68760
space technologies	0.03731
petroleum resources	-0.03613
biofuel energy	-0.00852
offshore wind turbine	0.00253
failure analysis	0.00246
chemical safety	-0.00170
environmentalism	-0.00127
beryllium disease	0.00115

Five slowest-growing concepts:

Concept 13	
Eigenvalue	5200
-------	
estonia	0.83787
book reviews	0.32368
renewable energy research	0.24135
wind systems	-0.16343
Editorials	0.15030
Biofuel	0.13599
Biofuels	-0.13414
energy technology	0.11337
petroleum resources	-0.11149
space technologies	-0.10324

Concept 27	
Eigenvalue	443
-------	
offshore wind turbine	-0.98746
beryllium disease	-0.08906
chronic beryllium disease	-0.08840
failure analysis	-0.06386
chemical safety	0.04588
wind systems	0.04145
legal environment	0.02451
carbon cap-and-trade	-0.01247
renewable energy incentives	-0.00947
wind energy	0.00924

Concept 28	
Eigenvalue	377
-------	
legal environment	-0.73959
failure analysis	0.44897
beryllium disease	-0.31065
chronic beryllium disease	-0.30621
chemical safety	0.23383
renewable energy incentives	0.04604
microhydro	-0.04475
energy politics	0.02204
editorials	0.01850
offshore wind turbine	0.01842

Concept 34	
Eigenvalue	85
-------	
ignalina nuclear power plant	-0.99988
estonia	0.01308
biofuel energy	0.00555
energy politics	0.00453
legal environment	0.00226
wind energy	0.00222
wind systems	-0.00172
offshore wind turbine	-0.00110
testing phase	0.00103
energy technology	-0.00085

Concept 37	
Eigenvalue	0
-------	
energy-saving technical progress	0.99453
hybrid diesel systems	0.09332
centre of appropriate technology	0.04637
load powering	0.00578
renewable energy development strategies	0.00395
market adoptions	-0.00188
renewable energy generation cost	0.00072
biogas electricity conversion	0.00012
chronic beryllium disease	0.00000
beryllium disease	0.00000

The fastest-growing concept is somewhat surprising, as its main components, *space technologies* and *petroleum resources*, do not represent particularly new fields. The next-highest concept makes more sense, however, as *renewable energy incentives* is becoming an increasingly important issue.

Most of the slowly growing concepts appear quite reasonable. It is not surprising that *estonia*, *legal environment*, *offshore wind turbine*, and *ignalina nuclear power plant* are not growing particularly rapidly in the context of renewable energy. The slowest-growing concept is represented by *energy-saving technical progress*, and while we would expect that there truly is increasing technical progress in the area of saving energy, it need not be the case that the term itself is growing at a high rate.

5.7 Renewable energy case study wrap-up

With the application of our software to the field of renewable energy, we see some positive results. The most important step of our methodology is perhaps the initial keyword extraction, as further analysis cannot produce quality results with poorly chosen terms as input. While our list certainly contains some irrelevant terms, such as *1.27e15 j*, and some overly general ones, such as *australia*, on the whole, the list is comprised of a relatively large proportion of potentially useful terms.

The term growth rankings are also promising. We see, for example, that *biofuel*, *biofuel energy*, *biofuels*, and *biogas electricity* are all highly ranked terms. However, the list of high-growth rate terms also contains several that are not of interest to us; these include, for example, *articles* and *book reviews*. We find that with large-scale studies such as this, with more than 1,000 initial keywords, it can be difficult to avoid such noise terms.

The results produced by LSA are very difficult to interpret in this case. Most of the generated concepts do not correspond particularly well with one's intuitive understanding of the term interrelationships. Although the term list was first filtered by growth rate, and subsequently by term impact level and generality, the number of terms was still too large to generate clean, easily understood concepts. LSA seems to be more effective when used on a small number of terms.

Finally, when calculating growth rates of the concepts identified with LSA, we find that the fastest-growing concept consists primarily of *space technologies*, *petroleum resources*, and *biofuel energy*. While the appearance of *biofuel energy* is a positive sign, the concept is unfortunately a grouping of three largely unrelated terms.

The results from this case study are promising in some regards, and show weakness in others. More work is needed before our methodology can be fully effective in technology forecasting. There are examples in these findings where manual expert intervention could help increase the quality of the results. But the goal of the project is to perform these analyses in a fully automated manner, and we have adhered strictly to this goal.

Chapter 6: Conclusions

6.1 Discussion

This paper has presented methods to aid researchers in understanding the complexities of their field. Our software gathers and analyzes information publicly available on the Internet to illuminate the relationships between fields and subfields, and to identify which of these subfields show promising growth. Bibliometrics and Latent Semantic Analysis are used to achieve these goals in a fully automated manner.

Our methodology is applicable to any field of research, but for concreteness, a case study on the field of renewable energy has been presented. The results of this study are promising. We identified over one thousand keywords related to the term *renewable energy*, of which a significant number represent important subfields. Trends in publication counts were used to identify the terms which are increasingly dramatically in prevalence, while LSA was used to group related terms into coherent concepts, highlighting term interrelationships and allowing us to analyze groups of terms as a single unit.

6.2 Review of objectives

Chapter 1 outlined five key objectives of this project. They are revisited here, with brief notes of our findings and how the objectives were achieved.

- **Discover related technologies**

Keyword extraction is introduced in section 3.2, and the relevant case study results are presented in section 5.2. Related technologies are identified by abstracting the concept of a particular technology as a single term, and using online publication search engines to identify other relevant terms.

Many potentially interesting terms were identified that were both relevant and highly focused on a subfield of renewable energy. However, some irrelevant terms and some overly general terms were also included.

- **Calculate a numerical value for the growth rate of a technology**

Growth rate analysis is discussed in section 3.4, with the relevant case study results presented in section 5.4. A technology's rate of growth is a function of the time series of its publication counts. The efficacy of several different metrics has been investigated.

The normalized ratio and normalized exponential rate metrics tend to be the most effective for our purposes of identifying technologies in the early stages of their growth. Results are generally positive, but on occasion, we find that the hit count is a poor indication of prevalence, and produces misleading growth rate results.

- **Determine the relationships among a set of technologies**

Latent Semantic Analysis identifies the concepts underlying a set of terms using term co-occurrence information. Each concept is comprised of a weighted combination of terms, where the highly weighted terms of a particular concept are related to one another. LSA is discussed in Chapter 4, and related case study results in section 5.5.

We have found that it becomes difficult to make sense of LSA results as the number of terms increases. Data cleaning is useful, however, in identifying and removing those terms which are the most prone to distort the results.

- **Identify important, fast-growing technologies in which an expert may be interested**

Given a "seed" technology, our methodology can be combined to identify relevant, potentially high growth technologies. First, terms related to the given technology are accumulated, possibly using back-pointing or eigenvector centrality testing to ensure relevance. The growth rates of each of these terms are then calculated, and those with the largest rates can be presented to an expert for review. This process is discussed in Chapter 3, and the relevant case study results are presented in sections 5.1 through 5.4.

The results of this process are promising. We have identified such terms as *ecologically optimized transport systems*, *microhydro*, *nanotechnology*, *biofuel*, and *biogas electricity conversion*. However, some unwanted terms are inevitably acquired from the keyword extraction step, and these terms can occasionally show large hit count growth, even when the seed term is included in the query. This causes false positives, terms which are improperly labeled as interesting subfields of the seed topic.

- **Validate our methods of technology identification by testing on fast-growing technologies of the past**

In section 3.4.2, a methodology is discussed for back-testing growth rate forecasts. Terms are identified which are known to have shown large growth over an extended period of time. Then, using only information available at the early stages of this period, we observe whether our software could have predicted such growth.

Back-testing has shown that our growth rate metrics tend to be consistent; that is, terms which are identified by large initial growth rates tend to continue showing large growth. Notably, every growth rate measure with the exception of the difference metric correctly foresaw the large growth of *biofuel* over the past ten years.

6.3 Future work

Thus far, we see promising results, but improvements can be made. Specifically, the keyword extraction and LSA modules could benefit from further refinement.

Keyword extraction suffers from an inherent tradeoff between relevance and specificity. The parameters of our algorithms are one-dimensional in this regard. With Engineering Village, the user chooses between controlled (more relevant) and uncontrolled (more specific) keywords. And with

Scirus, the user chooses between a wide or narrow search tree, which tend to produce more relevant and more specific terms, respectively. We attempt to overcome this tradeoff by first aiming to gather terms with a high degree of specificity, and then filtering these terms for relevance using back-pointing or eigenvector centrality testing. However, in practice this tends to again result in a set of terms which are overly general.

Another concern with the current term extraction results is that even terms which are both relevant and specific often do not represent technological fields, but rather locations, events, or issues surrounding the field of research. For example, in our case study we find terms such as *western australia*, *challenge 2008*, *global warming*, and *renewable energy plan 2012*. While terms such as these may be useful for some purposes, our goal is instead to identify related fields of research.

Improvements to the keyword extraction methodology may be possible by widening the range of sources. We are currently restricted to publication search engines, but sources such as blogs and wikis may also be well-suited for this project.

There is also work to be done in the area of LSA. Its results are often surprising or difficult to interpret. Although we already perform some data cleaning, it is possible that more extensive and systematic preprocessing techniques could be beneficial. It may also be the case that we would see more useful results by considering other algorithms that also analyze term co-occurrence counts. In either case, increasing the reliability of our results for term relationship identification would be quite beneficial.

We are already seeing encouraging results from our overall methodology, and with improvements to these areas, we believe that our system can provide considerable practical value to researchers.

Appendix: Large Result Sets

A.1 Engineering Village term extraction

Compendex controlled terms, fifty articles:

- *aerospace industry*
- *agricultural wastes*
- *air pollution*
- *biofuels*
- *biogas*
- *biomass*
- *building codes*
- *buildings*
- *carbon dioxide*
- *china*
- *climate change*
- *coal deposits*
- *competition*
- *computer aided instruction*
- *computer networks*
- *conservation*
- *construction industry*
- *cooling*
- *cost effectiveness*
- *cost reduction*
- *costs*
- *curricula*
- *decision making*
- *desalination*
- *design*
- *developing countries*
- *distance education*
- *distributed parameter networks*
- *distribution of goods*
- *ecology*
- *economic and social effects*
- *economics*
- *education*
- *electric batteries*
- *electric currents*
- *electric generators*
- *electric industry*
- *electric power distribution*
- *electric power generation*
- *electric power plants*
- *electric power systems*
- *electric power transmission networks*
- *electric power utilization*
- *electric utilities*
- *electricity*
- *electrochemistry*
- *emission control*
- *employment*
- *energy conservation*
- *energy conversion*
- *energy efficiency*
- *energy management*
- *energy policy*
- *energy resources*
- *energy storage*
- *energy utilization*
- *engineering education*
- *engineering research*
- *engines*
- *environmental engineering*
- *environmental impact*
- *environmental impact assessments*
- *environmental protection*
- *environmental testing*
- *ethanol*
- *extraction*
- *fluid dynamics*
- *flywheels*
- *fossil fuel power plants*
- *fossil fuels*
- *fuel cells*
- *fuel economy*
- *fuel storage*
- *gas emissions*
- *gas industry*
- *gas supply*
- *gas turbines*
- *gases*
- *geographic information systems*
- *geographical regions*
- *geothermal energy*
- *global warming*
- *greenhouse effect*
- *greenhouse gases*
- *harmonic analysis*
- *hydraulic machinery*
- *hydraulic motors*
- *hydroelectric generators*
- *hydroelectric power*
- *hydrogen*
- *hydrogen production*
- *hydrogen storage*
- *industrial economics*
- *intelligent buildings*
- *international law*
- *international trade*
- *internet*
- *investments*
- *landforms*
- *laws and legislation*
- *linear programming*
- *magnetic flux*
- *magnetism*
- *marketing*
- *mathematical models*
- *natural gas*
- *natural resources*
- *nonmetals*
- *nuclear fuel reprocessing*
- *ocean currents*
- *ocean thermal energy conversion*
- *oceanography*
- *offshore oil well production*
- *offshore structures*
- *optimization*
- *parks*
- *personnel training*
- *photovoltaic cells*
- *photovoltaic effects*
- *planning*
- *pollution*
- *power electronics*
- *power generation*
- *power plants*
- *power quality*
- *power transmission*
- *project management*
- *public policy*
- *public utilities*
- *radar stations*
- *regional planning*
- *regulatory compliance*
- *reliability*

- *renewable energy resources*
- *research and development management*
- *reservoirs (water)*
- *rotors*
- *rural areas*
- *sales*
- *satellites*
- *search engines*
- *semiconductor devices*
- *sensitivity analysis*
- *social aspects*
- *societies and institutions*
- *solar cells*
- *solar collectors*
- *solar energy*
- *solar heating*
- *solar power generation*
- *solar power plants*
- *solid oxide fuel cells (sofc)*
- *space applications*
- *speed*
- *standby power systems*
- *statistical tests*
- *stochastic models*
- *strategic planning*
- *structural design*
- *students*
- *surveys*
- *sustainable development*
- *systems analysis*
- *targets*
- *teaching*
- *technological forecasting*
- *technology*
- *technology transfer*
- *thermoelectric equipment*
- *tidal power*
- *transportation*
- *turbines*
- *turbulent flow*
- *underwater acoustics*
- *wakes*
- *water filtration*
- *wind effects*
- *wind power*
- *wind turbines*
- *world wide web*

Compendex uncontrolled terms, fifty articles:

- *annual growth*
- *argentina*
- *best practices*
- *bio energies*
- *bio energy*
- *bio-energy systems*
- *biofuel energy*
- *biogas electricity conversion*
- *building designers*
- *building regulations*
- *built environments*
- *calculation tools*
- *carbon cap-and-trade*
- *carbon dioxide discharges*
- *case studies*
- *clean development mechanism (cdm)*
- *community*
- *comparative analysis*
- *conventional energy sources*
- *conventional power*
- *conversion systems*
- *critical loads*
- *desalinated water*
- *developed countries*
- *development processes*
- *diffusion of renewables*
- *distributed energy*
- *distributed power*
- *domestic buildings*
- *dynamic testing*
- *ecologically sustainable development*
- *economic developments*
- *economic incentive*
- *education and trainings*
- *educational materials*
- *eirev*
- *electric demands*
- *electric range*
- *electrical and computer engineering (ece)*
- *electrical power facility (epf)*
- *electrical power test*
- *electricity generation*
- *electricity production*
- *electricity sales*
- *electricity sector*
- *elsevier (co)*
- *emission control policy*
- *emissions (asphalt)*
- *energy assessment methods*
- *energy conservation measures*
- *energy demands*
- *energy development*
- *energy mix*
- *energy politics*
- *energy production*
- *energy sector*
- *energy security*
- *energy supply*
- *energy surplus*
- *energy utilities*
- *environmental costs*
- *environmental performance*
- *environmental policies*
- *environmental pollution*
- *environmental pollutions*
- *environmentally friendly alternatives*
- *european renewable energy council (erec)*
- *fossil fuel combustion*
- *fuel savings*
- *germany*
- *global security*
- *government subsidies*
- *grid systems*
- *grid-connected systems*
- *hybrid renewable energy systems*
- *hydrogen systems*
- *incentive measures*
- *incentive mechanism*
- *india*
- *installed capacity*
- *institutional support*
- *integration system*
- *international conferences*
- *international initiatives*
- *islands*

- *kirklees councils*
- *little research*
- *load data*
- *load powering*
- *local authorities*
- *local employments*
- *local government*
- *local processing*
- *marine renewable energy*
- *market adoptions*
- *market participants*
- *millennium development goals (mdg)*
- *multilevel search engines*
- *national economy*
- *national parks*
- *negative effects*
- *non fossil fuel obligation*
- *non-renewable energy*
- *offshore environment*
- *optimal renewable energy mathematical (orem) model*
- *optimal renewable energy model*
- *optimization model*
- *optimization models*
- *output-based allocation*
- *pakistan*
- *participation*
- *peak energies*
- *peak powers*
- *performance indicators*
- *planning phases*
- *policies and measures to enhance use of renewable energy sources*
- *policy incentives*
- *policy makers*
- *policy options*
- *potential resources*
- *power data*
- *power sectors*
- *power stations*
- *power systems*
- *present situation*
- *primary factors*
- *production systems*
- *project datums*
- *project leaders*
- *promotion of renewable energy sources*
- *public acceptance*
- *public participation*
- *public perceptions*
- *qualitative data*
- *questionnaire surveys*
- *refurbishment projects*
- *regional development*
- *renewable electricity generation*
- *renewable energies*
- *renewable energy*
- *renewable energy (re)*
- *renewable energy (re) technologies*
- *renewable energy activities*
- *renewable energy assessment*
- *renewable energy capacity*
- *renewable energy centres*
- *renewable energy commercialization*
- *renewable energy development*
- *renewable energy developments*
- *renewable energy education*
- *renewable energy generation*
- *renewable energy industry*
- *renewable energy investments*
- *renewable energy market*
- *renewable energy penetration*
- *renewable energy planning*
- *renewable energy policy*
- *renewable energy potential*
- *renewable energy program*
- *renewable energy projects*
- *renewable energy promotion law*
- *renewable energy sector*
- *renewable energy sources*
- *renewable energy sources (res)*
- *renewable energy standards*
- *renewable energy statistics*
- *renewable energy subsidy policy*
- *renewable energy supply (res)*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable energy technology*
- *renewable energy technology (ret)*
- *renewable energy technology research*
- *renewable energy utilization*
- *renewable facilites*
- *renewable portfolio standards*
- *renewables*
- *research and developments*
- *rural communities*
- *rural energy services*
- *satellite-based sensors*
- *scale-up*
- *scaling up*
- *semi-structured interviews*
- *small communities*
- *space technologies*
- *stand-alone power systems*
- *stingray tidal current generator*
- *stingray tidal stream generator*
- *support mechanisms*
- *sustainable buildings*
- *sustainable development*
- *sustainable energy*
- *tax revenue*
- *technology push*
- *tertiary education*
- *testing phase*
- *time-dependent*
- *time-dependent systems*
- *top-down*
- *total energies*
- *transient modelling*
- *underwater power generation*
- *urban settings*
- *wake effects*
- *washington*
- *water demands*
- *wind facilities*
- *wind plants*
- *windfarm market*
- *world leaders*

Inspec controlled terms, fifty articles:

- *air pollution control*
- *bioenergy conversion*
- *biofuel*
- *carbon compounds*
- *climatology*
- *commerce*
- *computer aided instruction*
- *cooling*
- *cost reduction*
- *design engineering*
- *design for environment*
- *diesel-electric generators*
- *distance learning*
- *distribution networks*
- *economics*
- *educational courses*
- *electrical products*
- *electricity supply industry*
- *energy conservation*
- *energy resources*
- *environmental economics*
- *environmental factors*
- *environmental management*
- *failure analysis*
- *finance*
- *fossil fuels*
- *fuel cells*
- *geothermal power*
- *government policies*
- *heat pumps*
- *hybrid power systems*
- *hydroelectric power*
- *incentive schemes*
- *information resources*
- *international collaboration*
- *internet*
- *investment*
- *law*
- *legislation*
- *lighting*
- *linear programming*
- *marketing*
- *mathematical analysis*
- *nuclear power stations*
- *optimisation*
- *photovoltaic power systems*
- *pollution*
- *power consumption*
- *power electronics*
- *power engineering education*
- *power generation control*
- *power generation economics*
- *power generation planning*
- *power grids*
- *power markets*
- *power stations*
- *power system planning*
- *power utilisation*
- *pricing*
- *product design*
- *project engineering*
- *public utilities*
- *pumps*
- *recursive estimation*
- *reliability*
- *renewable energy sources*
- *research and development*
- *research and development management*
- *research initiatives*
- *sensitivity analysis*
- *socio-economic effects*
- *solar absorber-convertors*
- *solar cells*
- *solar heating*
- *solar power*
- *space heating*
- *standardisation*
- *standards*
- *sustainable development*
- *tariffs*
- *transportation*
- *unemployment*
- *wind power*
- *wind power plants*

Inspec uncontrolled terms, fifty articles:

- *1.27e15 j*
- *aboriginal communities*
- *accelerated depreciation plan*
- *ad 2020-2021*
- *after sales service*
- *agricultural sector*
- *alternative energy source*
- *articles*
- *asean economic sector*
- *atmospheric carbon dioxide*
- *australia*
- *australian cooperative research center for renewable energy*
- *australian cooperative research centre for renewable energy research program*
- *australian renewable energy sector*
- *auxiliary service*
- *awareness generation*
- *bachelor degree program*
- *balanced energy-mix*
- *baltic states*
- *bangladesh*
- *belgium*
- *bibliometric measures*
- *bio-energy systems*
- *biofuel*
- *biogas plants*
- *biogas system*
- *biomass*
- *biomass briquettes*
- *biomass energy*
- *biomass energy sources*
- *biomass potential*
- *bioresources*
- *book reviews*

- *broad policy guidelines*
- *bureau of energy of ministry of economical affairs*
- *business plans*
- *carbon dioxide emissions*
- *central australia*
- *centre of appropriate technology*
- *challenge 2008*
- *china*
- *china-india cooperation*
- *co2*
- *co2 emission control policy*
- *coal-fired power plants*
- *coherent renewable energy policy*
- *collaborative initiatives*
- *commercial energy scene*
- *commercial risks*
- *commercial sector*
- *commercialization*
- *community attitudes*
- *competitive electricity supply market*
- *competitive power markets*
- *consumer products*
- *continuous energy supply*
- *conventional generation cost*
- *cooking*
- *cooling*
- *cost/efficiency ratio minimisation*
- *course structure*
- *crude oil supplies reduction*
- *customer choice programs*
- *decarbonised world*
- *decentralised development*
- *decentralized power systems penetration*
- *delphi study*
- *desertification problem*
- *design for sustainability*
- *developing countries*
- *developing country*
- *diesel generators*
- *distributed applications*
- *distributed energy generation*
- *distribution networks*
- *domestic cooking*
- *domestic renewable energy generators*
- *domestic technology manufacturing*
- *donor funding*
- *drudgery reduction*
- *dynamic life cycle assessment*
- *earth climate change*
- *ecodesign*
- *ecologically optimized transport systems*
- *ecologically sustainable development*
- *economic aspects*
- *economic comparisons*
- *economic incentives*
- *economic topic*
- *economical potential*
- *editorials*
- *electric loads*
- *electric power policy frameworks*
- *electrical engineering*
- *electricity consumption*
- *electricity grid*
- *electricity industry restructuring*
- *electricity market reforms*
- *electricity price*
- *electricity prices*
- *electricity suppliers*
- *electricity supply industry*
- *emerging renewable energy markets*
- *emissions characteristics*
- *employment factors*
- *employment growth*
- *employment opportunities*
- *energy certificate*
- *energy conservation efforts*
- *energy converters*
- *energy demand*
- *energy demand increase*
- *energy demand variations*
- *energy density utilization*
- *energy diversification*
- *energy economics*
- *energy efficiency*
- *energy generation*
- *energy intensive economy*
- *energy market*
- *energy performance regulations*
- *energy planning*
- *energy policies*
- *energy policy*
- *energy policy act*
- *energy potential*
- *energy production*
- *energy requirement*
- *energy shortages*
- *energy sources*
- *energy studies*
- *energy supply*
- *energy supply security considerations*
- *energy systems engineering*
- *energy technology*
- *energy-price signals reformation*
- *engineering program*
- *enhanced greenhouse effect*
- *environment issues*
- *environmental contamination*
- *environmental cost*
- *environmental friendly sustainable development*
- *environmental impacts*
- *environmental interest*
- *environmental issues*
- *environmental pollution*
- *environmental problems*
- *environmental superiority*
- *environmentally friendly energy supply*
- *estonia*
- *eu energy performance of buildings directive*
- *european islands*
- *european union*
- *failure analysis*
- *federal commitment*
- *finance pre-feasibility studies*
- *financial assistance*
- *financial viability*
- *financing*
- *financing costs*
- *financing processes*
- *financing risk premiums reduction*
- *financing support*

- *finite energy resources*
- *flexible photo-voltaic solar cell*
- *fossil fuel*
- *fossil fuel economy*
- *fossil fuel electricity generation*
- *fossil fuel politics*
- *fossil fuels*
- *fuel cell*
- *fuel for thought strategies*
- *fuel standards*
- *fuel transportation*
- *funding mechanism*
- *future energy supply networks*
- *general public*
- *generation cost*
- *genoa*
- *geothermal energy*
- *geothermal heat pumps*
- *geothermal park*
- *german renewable energy source act*
- *germany*
- *global energy demand*
- *global renewable energy industry*
- *global renewable energy movement*
- *global warming*
- *government*
- *government incentives*
- *government mandates*
- *government of india*
- *government policies*
- *government policy decisions*
- *government subsidy*
- *government support*
- *greece*
- *green certificates*
- *green electricity*
- *green pricing*
- *greenhouse gas*
- *greenhouse gas reduction*
- *greenhouse gases*
- *greenhouse issues*
- *heat supply system*
- *heating*
- *high activity subjects*
- *high cost*
- *higher educational institutions*
- *hindu kush himalayan region*
- *household energy situation*
- *human health*
- *hybrid development*
- *hybrid diesel systems*
- *hybrid power systems*
- *hydro power*
- *hydroelectric power*
- *hydroelectricity*
- *hydrogen infrastructure*
- *ignalina nuclear power plant*
- *india*
- *indian national programme*
- *indian renewable energy development agency*
- *indian renewable energy development agency limited*
- *industrial sector*
- *innovative business models*
- *institutional barriers*
- *intermittent renewable energy penetration*
- *international cooperation*
- *international initiatives*
- *international renewable energy development*
- *internationally focused*
- *internet*
- *intrinsic technology problem*
- *investment*
- *investment decisions*
- *investment deduction*
- *investment incentives*
- *investments financing*
- *ipcc*
- *japan*
- *job losses*
- *joint ventures*
- *large scale dissemination*
- *latvia*
- *lca*
- *legal barriers*
- *legal environment*
- *legislation*
- *legislative aspects*
- *letters*
- *levelized energy costs*
- *lighting*
- *lighting end-uses*
- *lithuanian energy sector*
- *long-term stability*
- *low activity subjects*
- *low energy density*
- *low-interest loans*
- *maintenance*
- *malaysia*
- *marginalized mountain population*
- *market restructuring*
- *mechanical engineering*
- *mexico*
- *micro hydropower*
- *ministry of economic affairs of taiwan*
- *ministry of non-conventional energy sources*
- *ministry of nonconventional energy sources*
- *modelling study*
- *multidisciplinary programs*
- *murdoch university*
- *national energy planning*
- *national power installation capacity*
- *natural ecosystems*
- *natural gas consumption*
- *natural gas supplies reduction*
- *negative secondary impacts*
- *nepal*
- *nongovernmental organizations*
- *nonpooled electricity trading*
- *northern australia*
- *northern queensland*
- *northern south australia*
- *northern territory*
- *notes*
- *oil*
- *opec effect*
- *optimal renewable energy mathematical model*
- *optimal renewable energy model*
- *optimization model*
- *organizational aspects*
- *ownership system*
- *pakistan*
- *pastoral properties*

- *petroleum crude oil consumption*
- *photovoltaic energy*
- *photovoltaic generation system*
- *photovoltaic technology*
- *planning*
- *policies*
- *policies formulation*
- *policy design impacts*
- *policy makers*
- *policy making*
- *postgraduate program*
- *postgraduate university courses*
- *power electronics*
- *power grid interconnection*
- *power market*
- *power plant financing process*
- *power supply*
- *power system control*
- *power system design*
- *power system operation*
- *power system planning*
- *power systems engineering*
- *power utility*
- *price drops*
- *price mechanism*
- *private sector participation*
- *product design*
- *progressive utilization*
- *project development*
- *projects*
- *promotion strategies*
- *publications*
- *pumping*
- *pv*
- *pv power systems*
- *pv-compact*
- *quality control*
- *quality of life*
- *quota-based systems*
- *r&d*
- *r&d support*
- *rd and d*
- *recursive linear planning model*
- *recursive linear program*
- *regional power market competition*
- *regulated rural energy concessions*
- *regulatory law framework*
- *reliability*
- *reliability constraints*
- *remote area power supplies*
- *remote areas*
- *renewable energy*
- *renewable energy company limited*
- *renewable energy costs*
- *renewable energy curriculum*
- *renewable energy development*
- *renewable energy development bill*
- *renewable energy development plan*
- *renewable energy development status*
- *renewable energy development strategies*
- *renewable energy distribution*
- *renewable energy education*
- *renewable energy engineering curriculum*
- *renewable energy engineering system*
- *renewable energy equipment procurement*
- *renewable energy financing*
- *renewable energy generating equipments*
- *renewable energy generation*
- *renewable energy generation cost*
- *renewable energy incentives*
- *renewable energy industry*
- *renewable energy investment*
- *renewable energy law*
- *renewable energy literature*
- *renewable energy markets*
- *renewable energy materials*
- *renewable energy plan*
- *renewable energy plan 2012*
- *renewable energy planning*
- *renewable energy policies*
- *renewable energy policy*
- *renewable energy policy harmonisation*
- *renewable energy policy options*
- *renewable energy potential*
- *renewable energy product purchase*
- *renewable energy programmes*
- *renewable energy projects*
- *renewable energy promotion law*
- *renewable energy promotion strategies*
- *renewable energy quota system*
- *renewable energy r&d*
- *renewable energy research*
- *renewable energy resources*
- *renewable energy scenario*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy systems allocation*
- *renewable energy systems cost*
- *renewable energy systems reliability*
- *renewable energy technologies*
- *renewable energy technology*
- *renewable energy trading experience*
- *renewable energy usage*
- *renewable energy utilisation*
- *renewable energy utilization*
- *renewable portfolio*
- *renewislands*
- *research and development*
- *research efforts*
- *research output*
- *research priorities*
- *restructured electricity market*
- *revenue certainty*
- *reviews*
- *ring-fenced markets*
- *rural areas*
- *rural electrification*
- *rural energy enterprises*

- *sensitivity analysis*
- *small-scale renewable energy technologies*
- *social acceptance*
- *social acceptance variation*
- *social issues*
- *socio economic optimal renewable energy model*
- *solar city*
- *solar energy*
- *solar energy systems*
- *solar heat pump combination*
- *solar insolation*
- *solar photovoltaic*
- *solar photovoltaics*
- *solar pv*
- *solar pv utilization*
- *solar regime*
- *solar systems*
- *solar thermal energy*
- *spain*
- *standardization*
- *state government agencies*
- *statute for upgrading industries*
- *storage medium*
- *subsidy measures*
- *substitute cost*
- *supply shortages*
- *support mechanisms*
- *sustainable development*
- *sustainable energy*
- *sustainable energy source*
- *sustainable energy sources*
- *sustainable growth*
- *sustained renewable energy deployment*
- *taiwan*
- *tax credit*
- *tax revenue method*
- *technical potential*
- *technological aspects*
- *technological development*
- *tertiary renewable energy education*
- *tertiary trained engineers*
- *tertiary trained policy makers*
- *tertiary trained scientists*
- *time 2020 year to 2021 year*
- *total energy demand*
- *tourist ventures*
- *transmission networks*
- *transportation*
- *transportation sector*
- *turkey*
- *u.s. energy demand*
- *uk*
- *uk nffo*
- *undergraduate education*
- *undergraduate university courses*
- *unemployment rates*
- *uniform green power market mode*
- *urban areas*
- *usa*
- *waste energy resources*
- *web-based renewable energy courses*
- *western australia*
- *wind energy*
- *wind farm*
- *wind power*
- *wind power generation*
- *wind systems*
- *workshop in a box program*
- *world bank*
- *world wide web*

A.2 Scirus term extraction

Scirus terms, [10] pattern:

- *450*
- *access key*
- *active power*
- *actuators*
- *aerodynamic*
- *aerodynamics*
- *alarm limit*
- *alloy surfaces*
- *alternative energy*
- *alternative fuels*
- *american council*
- *amortization*
- *annual mean*
- *ashrae*
- *atmospheric sciences*
- *authorised*
- *base load*
- *battery consortium*
- *beryllium disease*
- *biodiesel*
- *biodiesel production*
- *bioenergy*
- *biofuel*
- *biogas*
- *biological attack*
- *biomass*
- *biomass combustion*
- *biomass energy*
- *biomass fuels*
- *biomass gasification*
- *biosphere*
- *blades*
- *borehole*
- *breeding populations*
- *broad-winged hawk*
- *building energy*
- *built environment*
- *business and industry*
- *capital cost*
- *carbon emissions*
- *carbon sequestration*
- *cash flow analysis*
- *center stage*
- *chronic beryllium disease*
- *clean coal*
- *climate control*
- *climate policy*
- *climatology*
- *co-operative*
- *co-ordination*
- *coalbed methane*
- *commodity standard*
- *computational science*
- *construction materials*
- *consumer behaviour*
- *contact dr*
- *conversion systems*
- *correlation*
- *cost reduction*
- *create jobs*
- *cross-flow*
- *data center*
- *de janeiro*
- *decision support system*
- *decision-support system*
- *delphi study*
- *deregulation*
- *development mechanism*
- *diesel*
- *digester gas*
- *dinosaurs*
- *distance learning*
- *distributed generation*
- *district heating*
- *doe program*
- *domestic energy*
- *domestic production*
- *donate*
- *drilled*
- *drilling operations*
- *dual fuel engine*
- *earth science*
- *economic costs*
- *economic dynamics*
- *economic information*
- *economic performance*
- *economics*
- *education programs*
- *efficiency measures*
- *efficient appliances*
- *electric energy*
- *electric grid*
- *electric power systems*
- *electric system*
- *electric utilities*
- *electricity consumption*
- *electricity demand*
- *electricity generation*
- *electricity generator*
- *electricity grid*
- *electricity industry*
- *electricity markets*
- *electricity prices*
- *electricity production*
- *electricity sector*
- *electricity supply*
- *electricity use*
- *electrification*
- *electrolysis*
- *electroplating*
- *emissions trading*
- *end-use efficiency*
- *energy*
- *energy audit*
- *energy capacity*
- *energy crisis*
- *energy crops*
- *energy development*
- *energy economics and policy*
- *energy efficiency*
- *energy farming*
- *energy impacts*
- *energy industry*
- *energy management*
- *energy output*
- *energy policy*
- *energy potential*
- *energy prices*
- *energy products*
- *energy research*
- *energy reserves*
- *energy resource*
- *energy resources*
- *energy savings*
- *energy sector*
- *energy services*
- *energy supplies*
- *energy systems*
- *energy technologies*
- *energy technology*
- *energy usage*
- *energy utilization*
- *energy-efficient lighting*

- *enthalpy*
- *environmental benefits*
- *environmental conservation*
- *environmental costs*
- *environmental education*
- *environmental protection agency*
- *environmental systems*
- *ethanol*
- *executive director*
- *exergy*
- *exergy efficiency*
- *experience curves*
- *fall in price*
- *farm operation*
- *fatigue life*
- *fault detection*
- *federal agencies*
- *federal reserve*
- *financial incentives*
- *form of energy*
- *fort collins*
- *fractures*
- *franchise tax*
- *fuel cell*
- *fuel cells*
- *fuel charge*
- *fuel economy*
- *fuel poverty*
- *future energy*
- *future generations*
- *galapagos archipelago*
- *gas exploration*
- *gas reserves*
- *gas-fired*
- *gasification*
- *gasoline*
- *gathering information*
- *gearbox*
- *general electric*
- *geoscience*
- *geothermal*
- *geothermal activity*
- *geothermal area*
- *geothermal areas*
- *geothermal development*
- *geothermal energy*
- *geothermal exploration*
- *geothermal field*
- *geothermal fields*
- *geothermal fluids*
- *geothermal gradient*
- *geothermal power*
- *geothermal reservoir*
- *geothermal reservoirs*
- *geothermal resources*
- *geothermal sources*
- *geothermal steam*
- *geothermal system*
- *geothermal systems*
- *geothermal water*
- *geothermal waters*
- *geothermal wells*
- *geyser*
- *gibraltar strait*
- *global climate change*
- *global markets*
- *global population*
- *golden eagles*
- *government entities*
- *great plains*
- *greenhouse gas abatement*
- *greenhouse gas emissions*
- *greenhouse heating*
- *greening*
- *ground loop*
- *grounding*
- *growth trends*
- *harmful emissions*
- *harnessing*
- *heat pump*
- *heat pump system*
- *heat pumps*
- *heating system*
- *heating systems*
- *high-altitude*
- *hot spring*
- *hot springs*
- *hydraulic*
- *hydro power*
- *hydroelectric*
- *hydroelectric dams*
- *hydrogen*
- *hydrogen fuel*
- *hydrogen transport*
- *hydrology*
- *hydropower*
- *hydrothermal*
- *impact on the environment*
- *improved energy*
- *incubator*
- *industrial ecology*
- *industrial energy*
- *industrial technologies*
- *information center*
- *information resources*
- *installed capacity*
- *institutional structures*
- *insulation*
- *integrated assessment*
- *integrated resource*
- *intelligent software*
- *intermittent*
- *international agreements*
- *international energy agency*
- *international law*
- *international network*
- *investor*
- *irrigation water pumping*
- *kilowatts*
- *land management*
- *landfill gas*
- *lead-acid batteries*
- *life cycle assessment*
- *life-cycle*
- *liquid column*
- *liquid dampers*
- *local residents*
- *long-term contracts*
- *low carbon*
- *low-income families*
- *market barriers*
- *mechanical power*
- *mechatronics*
- *megawatts*
- *metal recovery*
- *methane*
- *miles per hour*
- *milieu*
- *milieukunde*
- *mineral resources*
- *mitigation*
- *municipal utilities*
- *nameplate*
- *nanotechnology*
- *national security*
- *new zealand*
- *non-governmental organizations*
- *non-renewable*
- *novel design*
- *ocean breezes*
- *ocean thermal energy*
- *office of basic energy sciences*
- *office of science*
- *offshore wind turbine*
- *oil crisis*
- *oil prices*

- *oil reserves*
- *oil supply*
- *onshore*
- *open loop control system*
- *optimal placement*
- *optimization*
- *option to purchase*
- *organizational development*
- *outer continental shelf*
- *pelton wheel*
- *performance prediction*
- *petroleum reserves*
- *photovoltaic*
- *photovoltaic systems*
- *photovoltaics*
- *physical education*
- *pitch angle*
- *polluting*
- *poultry litter*
- *power coefficient*
- *power demand*
- *power development*
- *power electronics*
- *power generation*
- *power produced*
- *power production*
- *power quality*
- *power sources*
- *pressure decline*
- *producer gas*
- *production tax*
- *propulsion*
- *public affairs*
- *public benefits*
- *pump system*
- *pump systems*
- *rated power*
- *reactive power*
- *reducing greenhouse gas emissions*
- *reduction of carbon dioxide*
- *regulating*
- *reliability*
- *renewable energies*
- *renewable energy*
- *renewable energy resources*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable resource*
- *renewable sources*
- *renewables*
- *republic of croatia*
- *reservoir simulation*
- *residential buildings*
- *resource*
- *resource assessment*
- *resource potential*
- *resources management*
- *ress*
- *rhyolite*
- *rinse water*
- *roaring forties*
- *role of particles*
- *role of technology*
- *rotor blades*
- *rotor diameter*
- *rotor speed*
- *rural electrification*
- *rural electrification in senegal*
- *rural energy*
- *rural livelihoods*
- *salton trough*
- *sandstone*
- *savings plan*
- *scheikunde*
- *secretary of state*
- *semi-empirical models*
- *senegal*
- *sensitivity analysis*
- *service provider*
- *siting*
- *small island developing states*
- *solar cooker*
- *solar electricity*
- *solar energy conversion*
- *solar energy utilization*
- *solar photovoltaics*
- *solar power*
- *solar water*
- *solar water heaters*
- *solid oxide fuel cells*
- *sources of energy*
- *space heating*
- *spring water*
- *stand-alone*
- *state energy*
- *steam production*
- *storage technologies*
- *storage technology*
- *storm damage*
- *strait*
- *strong winds*
- *structural materials*
- *subcontractor*
- *subsurface*
- *sustainable development*
- *sustainable energy*
- *sustainable future*
- *system voltage*
- *tariff*
- *tasmania*
- *tax credit*
- *tax incentives*
- *taxpayer*
- *technological learning*
- *technology choice*
- *technology transfer*
- *thermal power*
- *thermal spring*
- *thermal springs*
- *thermal systems*
- *thermal water*
- *thermal waters*
- *thermochemical conversion*
- *tidal basin*
- *tidal energy*
- *tidal power*
- *tlcd*
- *transmission losses*
- *transport costs*
- *transportation fuels*
- *turbine*
- *turbine blades*
- *turbine rotor*
- *turbines*
- *turbulent*
- *turkey*
- *united nations development program*
- *united nations environment programme*
- *uranium*
- *uranium mining*
- *vibration control*
- *virgin islands*
- *volcanic zone*
- *wairakei*
- *wairakei field*
- *warm springs*
- *waste heat*
- *water conservation*
- *water heating*
- *water pumping*
- *water turbine*
- *wave energy*

- *wave loading*
- *welding*
- *west texas*
- *wind energy*
- *wind energy conversion*
- *wind energy converter*
- *wind energy potential*
- *wind events*
- *wind excitation*
- *wind farms*
- *wind generation*
- *wind loading*
- *wind power*
- *wind power development*
- *wind power plant*
- *wind resource*
- *wind speed data*
- *wind system*
- *wind systems*
- *wind turbine*
- *wind turbines*
- *wind velocity*
- *wood fiber*
- *wood waste*
- *world bank*

Scirus terms, [2, 2, 2] pattern:

- *450*
- *accelerator*
- *access key*
- *active power*
- *actuators*
- *advisory committee*
- *aerodynamic*
- *aerodynamics*
- *alarm limit*
- *alloy surfaces*
- *alternative energy*
- *alternative fuel*
- *alternative fuels*
- *american council*
- *amorphous silicon solar cells*
- *applied energy*
- *aquifer*
- *aquifers*
- *ashrae*
- *atmospheric sciences*
- *base load*
- *batch reactors*
- *battery consortium*
- *beryllium disease*
- *binary cycle*
- *biodiesel*
- *bioenergy*
- *biofuel*
- *biogas*
- *biological attack*
- *biomass*
- *biomass combustion*
- *biomass energy*
- *biomass fuels*
- *biomass gasification*
- *bioresource technology*
- *blades*
- *borehole*
- *breeding populations*
- *broad-winged hawk*
- *building energy*
- *built environment*
- *capital cost*
- *carbon emissions*
- *carbon sequestration*
- *chemical safety*
- *chronic beryllium disease*
- *clean coal*
- *clean fuels*
- *climate control*
- *climate policy*
- *climatology*
- *co-operative*
- *co-ordination*
- *coal liquefaction*
- *coal research*
- *computational science*
- *construction materials*
- *consumer behaviour*
- *contact dr*
- *conversion systems*
- *correlation*
- *cost reduction*
- *create jobs*
- *cross-flow*
- *data center*
- *de janeiro*
- *decision-making*
- *decision-support system*
- *deep drilling*
- *delphi study*
- *development mechanism*
- *diesel*
- *digester gas*
- *dinosaurs*
- *direct use*
- *distance learning*
- *distributed generation*
- *distribution of electricity*
- *district heating*
- *doe program*
- *domestic energy*
- *domestic production*
- *donate*
- *drilled*
- *drilling operations*
- *dry steam*
- *dual fuel engine*
- *earth science*
- *economic co-operation*
- *economic costs*
- *economic information*
- *economic performance*
- *education programs*
- *efficiency*
- *efficiency measures*
- *efficient appliances*
- *electric energy*
- *electric grid*
- *electric power systems*
- *electricity consumption*
- *electricity demand*
- *electricity generation*
- *electricity generator*
- *electricity grid*
- *electricity markets*
- *electricity production*
- *electricity sector*
- *electricity supply*
- *electricity use*
- *electrification*
- *electroplating*
- *emissions trading*
- *end-use efficiency*
- *energy*
- *energy applications*

- *energy capacity*
- *energy crops*
- *energy development*
- *energy economics*
- *energy efficiency*
- *energy farming*
- *energy impacts*
- *energy industry*
- *energy management*
- *energy output*
- *energy planning*
- *energy policy*
- *energy potential*
- *energy prices*
- *energy products*
- *energy research*
- *energy resource*
- *energy resources*
- *energy savings*
- *energy sector*
- *energy services*
- *energy supplies*
- *energy systems*
- *energy technologies*
- *energy technology*
- *energy usage*
- *energy utilization*
- *energy-efficient lighting*
- *enthalpy*
- *environmental benefits*
- *environmental conservation*
- *environmental costs*
- *environmental education*
- *environmental engineering*
- *environmental policy*
- *environmental protection agency*
- *environmental systems*
- *environmental technology*
- *environmentalism*
- *ethanol*
- *executive director*
- *exergy*
- *exergy analysis*
- *exergy efficiency*
- *experience curves*
- *fahrenheit*
- *fall in price*
- *farm operation*
- *fatigue life*
- *fault detection*
- *federal agencies*
- *financial assistance*
- *financial incentives*
- *fiscal year*
- *fission*
- *food technology*
- *form of energy*
- *fort collins*
- *fractures*
- *franchise tax*
- *freedom of information*
- *fuel cell*
- *fuel cell systems*
- *fuel cells*
- *fuel charge*
- *fuel cycle*
- *fuel economy*
- *fuel poverty*
- *fusion energy*
- *future energy*
- *future generations*
- *galapagos archipelago*
- *gas reserves*
- *gas supply*
- *gas-fired*
- *gas-tight*
- *gasoline*
- *gathering information*
- *gearbox*
- *general electric*
- *generation of electricity*
- *generation systems*
- *geoscience*
- *geothermal*
- *geothermal activity*
- *geothermal area*
- *geothermal areas*
- *geothermal development*
- *geothermal drilling*
- *geothermal energy*
- *geothermal exploration*
- *geothermal field*
- *geothermal fields*
- *geothermal fluids*
- *geothermal gradient*
- *geothermal plant*
- *geothermal power*
- *geothermal power plant*
- *geothermal power plants*
- *geothermal reservoir*
- *geothermal reservoirs*
- *geothermal resources*
- *geothermal resources -- environmental aspects -- oregon -- prineville region*
- *geothermal sources*
- *geothermal steam*
- *geothermal system*
- *geothermal systems*
- *geothermal water*
- *geothermal waters*
- *geothermal wells*
- *geyser*
- *gibraltar strait*
- *global climate change*
- *global issues*
- *global markets*
- *global population*
- *golden eagles*
- *government documents*
- *great basin*
- *great plains*
- *greenhouse gas abatement*
- *greenhouse gas emissions*
- *greenhouse heating*
- *ground loop*
- *grounding*
- *growing economy*
- *growth trends*
- *harmful emissions*
- *harnessing*
- *heat energy*
- *heat exchangers*
- *heat pump*
- *heat pump system*
- *heat pumps*
- *heated*
- *heating system*
- *heating systems*
- *high-altitude*
- *high-pressure*
- *hot spring*
- *hot springs*
- *human genome*
- *human genome project*
- *hydraulic*
- *hydro power*
- *hydroelectric*
- *hydroelectric dams*
- *hydrogen fuel*
- *hydrology*
- *hydropower*
- *hydrothermal*
- *hydrothermal systems*
- *impact on the environment*
- *incubator*
- *industrial ecology*
- *industrial energy*

- *industrial technologies*
- *information service*
- *innovation systems*
- *installed capacity*
- *insulation*
- *integrated assessment*
- *integrated gasification*
- *integrated resource*
- *intelligent software*
- *intergovernmental panel*
- *intergovernmental panel on climate change*
- *intermittent*
- *international agreements*
- *international energy agency*
- *international law*
- *international network*
- *investor*
- *irrigation water pumping*
- *issue date*
- *kilowatts*
- *land management*
- *landfill gas*
- *lead-acid batteries*
- *lead-acid battery*
- *life cycle assessment*
- *life-cycle*
- *liquefaction*
- *liquid column*
- *liquid dampers*
- *livermore national laboratory*
- *load factor*
- *local residents*
- *long-term contracts*
- *low carbon*
- *low-income families*
- *maps*
- *market barriers*
- *materials handling*
- *materials technology*
- *mechanical power*
- *mechatronics*
- *media relations*
- *megawatts*
- *metal recovery*
- *methane*
- *methane yield*
- *microforms*
- *miles per hour*
- *milieu*
- *milieukunde*
- *mitigation*
- *municipal utilities*
- *nameplate*
- *nanotechnology*
- *national security*
- *new zealand*
- *non-governmental organizations*
- *non-renewable*
- *novel design*
- *nuclear data*
- *nuclear research*
- *nuclear safety*
- *nuclear science*
- *nuclear technology*
- *nuclear weapons*
- *ocean breezes*
- *office of basic energy sciences*
- *office of energy research*
- *office of science*
- *offshore wind turbine*
- *oil crisis*
- *oil prices*
- *oil reserves*
- *oil supply*
- *onshore*
- *open loop control system*
- *optimal placement*
- *optimization*
- *option to purchase*
- *organisation for economic co-operation and development*
- *organizational development*
- *outlays*
- *pelton wheel*
- *performance prediction*
- *petroleum reserves*
- *petroleum resources*
- *photovoltaic*
- *photovoltaic cells*
- *photovoltaic systems*
- *photovoltaics*
- *physical education*
- *pitch angle*
- *plasma physics*
- *plate tectonics*
- *plutonium*
- *policy objectives*
- *pollutant emissions*
- *polluting*
- *poultry litter*
- *power coefficient*
- *power development*
- *power electronics*
- *power generation*
- *power produced*
- *power production*
- *power quality*
- *power sources*
- *pressure decline*
- *producer gas*
- *production tax*
- *propulsion*
- *public affairs*
- *public benefits*
- *public lands -- oregon -- prineville region -- management*
- *pump system*
- *pump systems*
- *rankine cycle*
- *rated power*
- *reactive power*
- *record search*
- *reducing greenhouse gas emissions*
- *reduction of carbon dioxide*
- *regulating*
- *reliability*
- *reliable approach*
- *renewable energies*
- *renewable energy*
- *renewable energy resources*
- *renewable energy sources*
- *renewable energy systems*
- *renewable energy technologies*
- *renewable resource*
- *renewable sources*
- *renewables*
- *replenish*
- *republic of croatia*
- *reservoir simulation*
- *residential buildings*
- *resource assessment*
- *resource potential*
- *ress*
- *rhyolite*
- *rinse water*
- *roaring forties*
- *rock formations*
- *role of particles*
- *role of technology*
- *rotor blades*
- *rotor diameter*

- *rotor speed*
- *rural electrification*
- *rural electrification in senegal*
- *rural energy*
- *rural livelihoods*
- *salton sea*
- *salton trough*
- *sandstone*
- *scheikunde*
- *secretary of state*
- *semi-empirical models*
- *senegal*
- *senior vice president*
- *sequencing batch*
- *sequencing batch reactors*
- *service provider*
- *short segment*
- *siting*
- *small island developing states*
- *social studies*
- *soil gas*
- *solar cell*
- *solar collectors*
- *solar cooker*
- *solar electricity*
- *solar energy applications*
- *solar energy conversion*
- *solar energy materials and solar cells*
- *solar energy utilization*
- *solar photovoltaics*
- *solar power*
- *solar water*
- *solar water heaters*
- *sources of energy*
- *space heating*
- *spring water*
- *stand-alone*
- *state energy*
- *state government*
- *state of michigan*
- *steam power*
- *steam production*
- *storage technologies*
- *storage technology*
- *storm damage*
- *strait*
- *strong winds*
- *subcontractor*
- *subsurface*
- *supercomputing*
- *sustainable development*
- *sustainable energy*
- *sustainable future*
- *swine manure*
- *system voltage*
- *systems research*
- *tariff*
- *tasmania*
- *tax credit*
- *tax incentives*
- *taxpayer*
- *technological learning*
- *technological research*
- *technology choice*
- *technology division*
- *technology transfer*
- *terrorism*
- *test reactor*
- *thermal power*
- *thermal spring*
- *thermal springs*
- *thermal systems*
- *thermal water*
- *thermal waters*
- *thermochemical conversion*
- *thermonuclear power*
- *tidal energy*
- *tidal power*
- *tlcd*
- *transportation fuels*
- *triple point*
- *turbine*
- *turbine blades*
- *turbine rotor*
- *turbines*
- *turbulent*
- *turkey*
- *united nations development program*
- *united nations environment programme*
- *uranium*
- *uranium mining*
- *vibration control*
- *vice president*
- *virgin islands*
- *volcanic zone*
- *volcano*
- *wairakei*
- *wairakei field*
- *warm springs*
- *waste heat*
- *water heaters*
- *water heating*
- *water pumping*
- *water turbine*
- *wave energy*
- *wave loading*
- *welding*
- *west texas*
- *western flank*
- *wind energy*
- *wind energy conversion*
- *wind energy converter*
- *wind energy potential*
- *wind events*
- *wind excitation*
- *wind farms*
- *wind generation*
- *wind loading*
- *wind power*
- *wind power development*
- *wind power plant*
- *wind resource*
- *wind speed data*
- *wind system*
- *wind systems*
- *wind turbine*
- *wind turbines*
- *wind velocity*
- *wood waste*
- *working fluid*
- *world bank*

A.3 Back-pointing relevance testing

Term generation by back-pointing, 1,000 potential terms, three levels:

Level 1

- *biomass fuels*
- *wind power development*
- *photovoltaic system*
- *turbines*
- *geothermal resources*
- *experience curves*
- *electrification*
- *exergy efficiency*
- *biomass conversion*
- *geothermal systems*
- *small wind turbines*
- *turbine*
- *energy situation*
- *microhydro*
- *electricity production*
- *biomass resources*
- *low wind speed turbine*
- *kilowatts*
- *novel design*
- *wind energy converter*
- *renewable sources*
- *renewable system*
- *transport sector*
- *megawatts*
- *installed power*
- *hydroelectric*
- *hydropower*
- *water heating*
- *biogas*
- *geothermal development*
- *wave energy*
- *solar power plant*
- *wind energy potential*
- *turbine generator*
- *energy planning*
- *energy supplies*
- *wind farm*
- *emissions reduction*
- *business reasons*
- *renewable energies*
- *rural energy*
- *tidal stream*
- *energy impacts*
- *wind turbines*
- *low head*
- *tidal power*
- *energy crops*
- *rural electrification*
- *power development*
- *remote power generation*
- *wind farms*
- *geothermal sources*
- *wind*
- *wind turbine*

Level 2

- *alternator*
- *industrial gas turbine*
- *ice accretion*
- *electric motor*
- *flowmeter*
- *energy performance*
- *thermochemical*
- *electricity*
- *energy industry*
- *short rotation*
- *photovoltaic cell*
- *power rating*
- *distributed generation*
- *technological learning*
- *power quality*
- *gearbox*
- *geothermal power plant*
- *renewable energy sources*
- *methane*
- *turbine blade*
- *power take-off*
- *exergy*
- *biomass*
- *center stage*
- *heat pump system*
- *offshore environment*
- *solar energy applications*
- *wairakei*
- *biomass energy*
- *amount of energy*
- *alternative energy*
- *crop residues*
- *kyoto protocol*
- *flash steam*
- *heat pumps*
- *energy resources*
- *greenhouse gas emissions*
- *energy potential*
- *turbine engine*
- *electric system*
- *water heaters*
- *aerodynamic*
- *environmental benefits*
- *eigenvalue analysis*
- *energy efficient*
- *digester gas*
- *storage technologies*
- *state energy*
- *decision support system*
- *blade root*
- *power coefficient*
- *landfill gas*
- *geothermal waters*
- *non-polluting*
- *sensitivity analysis*
- *battery storage*
- *renewable resource*
- *sustainable energy*
- *sources of energy*
- *battery backup*
- *irrigation water pumping*
- *hydrogen fuel*
- *municipal utilities*
- *rocket engines*
- *performance prediction*
- *energy efficiency*
- *solar heating*
- *harnessing*
- *electricity generator*
- *life cycle assessment*
- *electrical network*
- *energy prices*

- *rotor speed*
- *electricity prices*
- *hydrogen*
- *energy systems*
- *heating systems*
- *district heating*
- *geothermal fluids*
- *tidal basin*
- *excess power*
- *thermal water*
- *energy technology*
- *environmental affairs*
- *production tax*
- *generating units*
- *domestic production*
- *intermediate pressure*
- *electricity markets*
- *specific speed*
- *dry steam*
- *steam turbine*
- *mechanical power*
- *turbine blades*
- *geothermal energy*
- *fuel cell*
- *geothermal steam*
- *future energy*
- *tidal energy*
- *solar energy utilization*
- *power produced*
- *domestic hot water*
- *solar systems*
- *flat terrain*
- *emissions reductions*
- *greenhouse heating*
- *aerodynamics*
- *hydroelectric dams*
- *steel grade*
- *solid oxide fuel cells*
- *combustion technology*
- *energy economics*
- *non-renewable*
- *gas turbine engine*
- *wind power plant*
- *diesel*
- *particle erosion*
- *solar collector*
- *geothermal power plants*
- *brewery*
- *impulse turbine*
- *guide vanes*
- *electricity use*
- *solar energy systems*
- *solar collectors*
- *emissions trading*
- *energy services*
- *energy development*
- *thermal power*
- *energy resource*
- *rotor blade*
- *geothermal system*
- *power sector*
- *rotor blades*
- *salton trough*
- *energy conservation*
- *intelligent software*
- *hydro power*
- *film cooling*
- *high temperature fatigue*
- *condition monitoring*
- *emissions inventory*
- *wind energy*
- *loan guarantee*
- *international emissions*
- *polarization curve*
- *policy instruments*
- *heat energy*
- *wind systems*
- *solar photovoltaics*
- *generator system*
- *wind power*
- *geothermal field*
- *fermentation*
- *integrated gasification*
- *combustion chamber*
- *roaring forties*
- *plant design*
- *tc hydraulic engineering*
- *ocean engineering*
- *reduction of carbon dioxide*
- *energy savings*
- *combustors*
- *solar generator*
- *germany*
- *gas field*
- *electricity generation*
- *electric power systems*
- *solar water*
- *propulsion*
- *electricity supply industry*
- *electricity demand*
- *coal-fired power plants*
- *dynamic stall*
- *emissions permits*
- *industrial gas*
- *planetary gear*
- *wind engineering*
- *geothermal water*
- *generating station*
- *feedstock*
- *water turbine*
- *renewable energy systems*
- *energy recovery*
- *industrial energy*
- *steam turbines*
- *rotor diameter*
- *emission factors*
- *geothermal gradient*
- *power equipment*
- *water pumping*
- *generation of electricity*
- *horizontal axis*
- *conversion systems*
- *gas turbine engines*
- *gas turbine*
- *energy capacity*
- *carbon emissions*
- *renewable energy technologies*
- *atmospheric concentrations*
- *biofuel*
- *emission reduction*
- *light energy*
- *scheikunde*
- *solar cooker*
- *geothermal reservoirs*
- *solar hydrogen*
- *wind energy conversion*
- *geothermal activity*
- *portfolio approach*
- *utilization factor*
- *photovoltaics*
- *rated power*
- *heat pump*
- *photovoltaic power*
- *deregulation*
- *solar power generation*
- *electric generator*
- *space heating*
- *harmful emissions*
- *power demand*
- *wairakei field*
- *biomass combustion*
- *steam power*
- *geothermal fields*
- *thermal systems*
- *bioelectric energy sources*
- *environmental impact assessment*
- *electricity consumption*

- *jet engine*
- *renewables*
- *siting*
- *solar cookers*
- *electricity distribution*
- *distribution of electricity*
- *geothermal*
- *maximum power density*
- *market barriers*
- *thermal wind*
- *wind resource*
- *geothermal plant*
- *geothermal wells*
- *solar house*
- *electricity grid*
- *turbine noise*
- *simple cycle*
- *base load*
- *energy utilization*
- *iron and steel industry*
- *end-use efficiency*
- *turbine rotor*
- *energy crisis*
- *improved energy*
- *wind generation*
- *emission reductions*
- *transportation sector*
- *general electric*
- *reactive power*
- *solar power*
- *electricity sector*
- *energy technologies*
- *sectoral*
- *electricity supply*
- *power generation*
- *gas-fired*
- *geothermal area*
- *golden eagles*
- *farm operation*
- *energy sector*
- *bioenergy*
- *ceramic matrix*
- *energy usage*
- *pitch angle*
- *solar electricity*
- *siemens*
- *waste heat*
- *lca*
- *thermochemical conversion*
- *energy policy*
- *electric grid*
- *nameplate*
- *cogeneration*
- *energy output*
- *taupo volcanic zone*
- *blades*
- *biogas plants*
- *carbon intensity*
- *ocean breezes*
- *energy crop*
- *coal-fired plants*
- *agricultural residues*
- *electricity industry*
- *clean coal*
- *gasification*
- *economical*
- *hot springs*
- *slip rings*
- *variable pitch*
- *greenhouse gas*
- *mass flow rate*
- *geothermal power*
- *renewable energy resources*
- *turbine engines*
- *solar heat*
- *wind system*
- *climate change policy*
- *electric utilities*
- *photovoltaic*
- *domestic energy*
- *volcanic zone*
- *onshore*
- *geothermal reservoir*
- *photovoltaic systems*
- *sustainable development*
- *angle of attack*
- *energy*
- *role of hydrogen*
- *fatigue loads*
- *cooling air*
- *exhaust system*
- *heating system*
- *wind events*
- *home energy*
- *wind speed data*
- *installed capacity*
- *efficient appliances*
- *bioelectricity*
- *power production*
- *space cooling*
- *desalination*
- *carbon dioxide emissions*
- *geothermal areas*
- *development mechanism*
- *collector system*
- *pv systems*
- *reliability*
- *air emissions*
- *gear box*
- *anaerobic*
- *nox emissions*
- *power sources*
- *geothermal exploration*
- *energy research*
- *biomass gasification*
- *reducing greenhouse gas emissions*
- *co2 emissions*
- *fuel efficiency*
- *heat exchangers*
- *energy audit*
- *electric energy*
- *energy system analysis*
- *alternative fuels*
- *pitch diameter*
- *low carbon*
- *energy products*
- *nature conservation*
- *capacity factor*
- *microbial fuel cell*
- *geothermal gradients*

Level 3

- *irradiance*
- *non-renewable resource*
- *extraction policy*
- *ceramic*
- *abatement*
- *wind velocity*
- *industrial park*
- *open-circuit voltage*
- *energy-efficient lighting*
- *trade group*
- *refocus*
- *municipal solid waste*

- *pump systems*
- *life-cycle*
- *heat storage*
- *chemical oxygen demand*
- *fuel economy*
- *mitigation measures*
- *west texas*
- *gas supply*
- *nonlinear control*
- *rotary motion*
- *rechargeable batteries*
- *semi-empirical models*
- *national wildlife refuge*
- *sewage*
- *creation of credit*
- *component testing*
- *global climate change*
- *central shaft*
- *balanced growth*
- *environmental justice*
- *economic cooperation*
- *small island developing states*
- *oil companies*
- *optimization*
- *oil prices*
- *polluting*
- *[jel] q3*
- *agricultural waste*
- *atmospheric carbon dioxide levels*
- *ethanol*
- *hydrofluorocarbons*
- *supplying*
- *electrolysis*
- *energy economics and policy*
- *pyrolysis*
- *economic dynamics*
- *atmospheric emissions*
- *gaseous fuel*
- *climate control*
- *hvac system*
- *computational science*
- *capital cost*
- *wilderness areas*
- *fuel charge*
- *full-sized*
- *new zealand*
- *construction materials*
- *usa*
- *natural forests*
- *original equipment manufacturers*
- *mass conservation*
- *regional economy*
- *pellet production*
- *tariff*
- *systems research*
- *wine industry*
- *launch vehicle*
- *metal recovery*
- *gasoline*
- *hydrolysis*
- *thermal storage*
- *drilling operations*
- *expected life*
- *enter into force*
- *oil drilling*
- *methanol*
- *biodiesel*
- *buildings*
- *financial incentives*
- *resource potential*
- *cost reduction*
- *savings plan*
- *reactor safety*
- *silicon wafers*
- *electronic components*
- *exhaustible resources*
- *control circuit*
- *role of particles*
- *voltage control*
- *environmental groups*
- *climate policy*
- *commercial buildings*
- *ignition system*
- *matrix composites*
- *labor input*
- *building services*
- *sulphur hexafluoride*
- *framework convention*
- *territorial*
- *environmental law*
- *bipartisan*
- *cell technology*
- *industrial ecology*
- *hot spring*
- *issue date*
- *cathodes*
- *seasonal demand*
- *radiant*
- *reformulated gasoline*
- *volcano*
- *developing countries*
- *thermal waters*
- *nitrous oxide emissions*
- *non-renewable resources*
- *waste land*
- *producer gas*
- *future generations*
- *internal resistance*
- *power pack*
- *form of energy*
- *impacts of climate change*
- *power electronics*
- *distribution networks*
- *rice hulls*
- *changing climate*
- *gas exploration*
- *local residents*
- *biodiesel production*
- *exhaustible*
- *united nations environment programme*
- *insulation*
- *chemical engineer*
- *organic substrate*
- *trading systems*
- *sustainable future*
- *organic solar cells*
- *forest biomass*
- *semi-endogenous growth*
- *economics association*
- *sulphur*
- *greenhouse gas abatement*
- *emission trading*
- *energy management*
- *international space station*
- *intergenerational equity*
- *rate of extraction*
- *fuel poverty*
- *borehole*
- *integrated resource*
- *ground loop*
- *business climate*
- *resource assessment*
- *resource depletion*
- *office of science*
- *electric power transmission*
- *pump system*
- *uranium mining*
- *solar energy conversion*
- *strong winds*
- *cooling system*
- *resource and energy economics*

- *environmental biotechnology*
- *energy intensity*
- *intergovernmental panel on climate change*
- *depletion rate*
- *nitrous oxide emission*
- *mechatronics*
- *external cost*
- *petroleum resources*
- *ventilation air*
- *central heating*
- *methane emissions*
- *solar cell*
- *office of basic energy sciences*
- *global markets*
- *vehicle fleet*
- *geyser*
- *asset management*
- *nitrous*
- *minnesota*
- *environmental economics*
- *environmental education*
- *economic systems*
- *transitional dynamics*
- *resource*
- *nuclear safety*
- *extracted resources*
- *endogenous technical change*
- *externalities*
- *electrodes*
- *photovoltaic cells*
- *mineral resources*
- *gas industry*
- *market mechanisms*
- *environmental costs*
- *impact on the environment*
- *nuclear industry*
- *fort lauderdale*
- *battery consortium*
- *paper company*
- *international energy agency*
- *fire protection association*
- *geosynchronous orbit*
- *external resistance*
- *coalbed methane*
- *efficiency*
- *executive director*
- *policy implementation*
- *frequency control*
- *rhyolite*
- *turbine flowmeter*
- *trading program*
- *use efficiency*
- *oil reserves*
- *policy recommendations*
- *economic effects*
- *rinse water*
- *biofilm*
- *drilled*
- *ecological footprint*
- *land-use*
- *sea trials*
- *industrial technologies*
- *environmental performance*
- *mitigation*
- *ignimbrite*
- *innovation policy*
- *hot-humid climate*
- *annual mean*
- *high-altitude*
- *gas turbines*
- *co-ordination*
- *enthalpy*
- *environmental systems*
- *spot market*
- *dual fuel engine*
- *welding*
- *stationary sources*
- *economic affairs*
- *environment agency*
- *guaranteed loan*
- *oil*
- *long-term contracts*
- *amortization*
- *government entities*
- *heat rejection*
- *oil crisis*
- *environmental engineering*
- *manifold*
- *sequestration*
- *fuel cells*
- *lead-acid batteries*
- *miles per hour*
- *balanced growth path*
- *market trends*
- *environmental sustainability*
- *active power*
- *public affairs*
- *price increase*
- *carbon sequestration*
- *military aircraft*
- *load factor*
- *gas fields*
- *residential buildings*
- *system voltage*
- *wood waste*
- *storage technology*
- *hybrid poplar*
- *nanotechnology*
- *gas reserves*
- *commodity standard*
- *cross flow*
- *o emissions*
- *steam production*
- *vortex*
- *resource extraction*
- *environment programme*
- *increasing returns*
- *vice president*
- *cathode*
- *ashrae*
- *biocatalyst*
- *global carbon*
- *poultry litter*
- *[jel] q32*
- *environmental conservation*
- *building energy*
- *improved model*
- *vibration control*
- *liquid fuel*
- *reverse osmosis*
- *conservation reserve program*
- *equilibrium path*
- *green taxes*
- *republic of croatia*
- *outer continental shelf*
- *tax credit*
- *optimal placement*
- *tax incentives*
- *national security*
- *steel production*
- *friction loss*
- *greening*
- *land management*
- *warm springs*
- *market simulation*
- *permafrost*
- *flue gas*
- *american council*
- *integrated assessment*
- *petroleum reserves*
- *susceptance*
- *environmental policy*
- *heated*

- *terminal voltage*
- *inventories*
- *institutional structures*
- *tax breaks*
- *power supplies*
- *heating oil*
- *ethanol production*
- *sustainable welfare*
- *built environment*
- *energy reserves*
- *role of technology*
- *hvac*
- *learning rates*
- *fault detection*
- *nonrenewable resources*
- *oil equivalent*
- *project manager*
- *hydrogen storage*
- *island press*
- *eastern european countries*
- *load management*
- *efficiency measures*
- *energy-saving technical progress*
- *transportation fuels*
- *maldives*
- *oil supply*
- *spring water*
- *atmospheric gases*
- *doe program*
- *national bureau of economic research*
- *effluents*

Term generation by back-pointing, 5,000 potential terms, one level, *renewable energy*:

- *biomass fuels*
- *photovoltaic system*
- *hydroelectric energy*
- *kizildere*
- *electrification*
- *aggregated wind turbines*
- *exergy efficiency*
- *small wind turbines*
- *turbine*
- *low wind speed turbine*
- *novel design*
- *wind energy converter*
- *hybrid solar-wind system*
- *megawatts*
- *energy applications*
- *hydroelectric*
- *wind atlas*
- *installed power*
- *solar power plant*
- *tidal barrage*
- *electricity production*
- *business reasons*
- *heating water*
- *kilowatts*
- *tidal power station*
- *rural electrification*
- *wind farms*
- *barrels per day*
- *ocean energy*
- *regional energy planning*
- *residential building*
- *[jel] q48*
- *microhydro*
- *renewable sources*
- *solar heating system*
- *geothermal development*
- *energy planning*
- *emissions reduction*
- *renewable energies*
- *photovoltaic array*
- *tidal stream*
- *wind turbines*
- *low head*
- *wind power development*
- *stream power*
- *remote power generation*
- *solar water heating systems*
- *geothermal resources*
- *experience curves*
- *wind energy conversion systems*
- *turkey*
- *electrolyzer*
- *renewable system*
- *transport sector*
- *hydropower*
- *exergy analysis*
- *solar water heating*
- *rural energy*
- *energy impacts*
- *energy crops*
- *wave direction*
- *linear generator*
- *tidal power*
- *wind power density*
- *geothermal systems*
- *ocean wave energy*
- *biomass resources*
- *significant wave height*
- *energy situation*
- *electrodialysis*
- *biomass conversion*
- *water heating*
- *biogas*
- *wave energy*
- *solar lanterns*
- *wind energy potential*
- *turbine generator*
- *wind energy utilization*
- *energy supplies*
- *wind farm*
- *wave energy conversion*
- *turbines*
- *power development*
- *wind*
- *tidal channel*
- *geothermal sources*
- *tidal current*
- *wind turbine*

Term generation by back-pointing, 5,000 potential terms, one level, *sustainable energy*:

- *renewable energy*
- *energy technologies*
- *renewable energy sources*
- *renewable energy technologies*
- *wind energy*
- *wind power*
- *renewable sources*
- *geothermal*

- *energy resources*
- *wind turbines*
- *solar power*
- *renewable energy resources*
- *turbines*
- *renewable energy systems*
- *energy systems*
- *renewable energies*
- *biomass energy*
- *energy policy*
- *energy industry*
- *geothermal energy*
- *electricity generation*
- *turbine*
- *wind farms*
- *alternative energy*
- *hydroelectric*
- *sources of energy*
- *energy development*
- *rural energy*
- *state energy*
- *energy research*
- *distributed generation*
- *wave energy*
- *energy technology*
- *energy crops*
- *wind resource*
- *energy services*
- *landfill gas*
- *rural electrification*
- *hydrogen fuel*
- *alternative fuels*
- *electrification*
- *tidal power*
- *solar photovoltaics*
- *tidal energy*
- *efficient appliances*
- *fuel cells*
- *future energy*
- *built environment*
- *hydro power*
- *electric energy*
- *water heating*
- *energy utilization*
- *solar electricity*
- *energy sector*
- *energy management*
- *energy-efficient lighting*
- *thermal systems*
- *energy products*
- *building energy*
- *domestic energy*
- *energy supplies*
- *storage technologies*
- *bioenergy*
- *installed capacity*
- *solar water*
- *photovoltaic systems*
- *electricity production*
- *energy impacts*
- *energy efficiency*
- *distribution of electricity*
- *biodiesel*
- *biomass*
- *sustainable future*
- *form of energy*
- *electricity sector*
- *market barriers*
- *irrigation water pumping*
- *energy resource*
- *small island developing states*
- *maldives*
- *wind*
- *wind systems*
- *wind energy potential*
- *megawatts*
- *energy capacity*
- *electricity supply*
- *kilowatts*
- *wind power development*
- *power development*
- *industrial energy*
- *electricity grid*
- *hydropower*
- *hydrogen*
- *energy economics and policy*
- *energy planning*
- *electricity*
- *institutional structures*
- *solar hydrogen*
- *geothermal resources*
- *geothermal development*
- *geothermal water*
- *geothermal sources*
- *geothermal fluids*
- *energy*
- *renewables*
- *base load*
- *resource potential*
- *energy crisis*
- *electricity industry*
- *energy audit*
- *global climate change*
- *international energy agency*
- *nanotechnology*
- *low carbon*
- *mechatronics*
- *computational science*
- *end-use efficiency*
- *united nations environment programme*
- *fuel poverty*
- *wind farm*
- *solar systems*
- *solar heating*
- *light energy*
- *solar house*
- *solar power plant*
- *solar energy applications*
- *cell technology*
- *electricity use*
- *energy efficient*
- *energy-saving technical progress*
- *tidal stream*
- *horizontal axis*
- *microhydro*
- *home energy*
- *remote power generation*
- *renewable system*
- *sustainable development*
- *energy system analysis*
- *energy conservation*
- *buildings*
- *usa*
- *energy economics*
- *solar energy systems*
- *load management*
- *nuclear safety*
- *role of hydrogen*
- *biomass resources*
- *emissions reduction*
- *energy situation*
- *power sector*
- *climate policy*
- *environment programme*
- *environmental groups*
- *deregulation*
- *fuel efficiency*
- *heat energy*
- *generation of electricity*
- *electricity supply industry*
- *small wind turbines*
- *environmental impact assessment*
- *installed power*

- *electric power systems research*
- *aggregated wind turbines*
- *rated capacity*
- *alternative fuel*
- *alternative energy source*
- *energy evaluation*
- *hydroelectric energy*
- *spatial planning*
- *licensing procedure*
- *solar wind*
- *fuel power*
- *electric vehicles*
- *efficient lighting*
- *united nations industrial development organization*
- *rural economy*
- *energy consumption*
- *energy model*
- *hydrogen energy*
- *solar energy*
- *building industry*
- *clean fuels*
- *applied energy*
- *policy reform*
- *combined heat and power*
- *energy extraction*
- *commercialisation*
- *wave power*
- *wave energy conversion*
- *fuel cell systems*
- *price of energy*
- *solar energy materials and solar cells*
- *energy conversion and management*
- *framework programme*
- *kizildere*
- *energy conversion systems*
- *competing technologies*
- *international bank for reconstruction and development*
- *local wind*
- *hybrid solar-wind system*
- *wind resource assessment*
- *power resource*
- *wind power density*
- *progress in energy and combustion science*
- *wind energy utilization*
- *wind atlas*
- *electric vehicle*
- *hydrogen vehicles*
- *fuel storage*
- *energy intensities*
- *methane generation*
- *energy unit*
- *use of hydrogen*
- *energy carrier*
- *production of hydrogen*
- *petroleum gas*
- *tidal power station*
- *tidal barrage*
- *sustainable use*
- *solar concentrator*
- *solar lanterns*
- *solar water heating systems*
- *energy conversion efficiency*
- *tidal wave*
- *tidal elevation*
- *market transformation*
- *hot water systems*
- *overview of fuel cell*
- *sustainable growth*
- *school environment*
- *fossil fuels*
- *biofuels*
- *energy-generating resources*
- *final energy*
- *hydro-electric power*
- *cascade of hydro power plants*
- *greenhouse climate control*
- *energy and buildings*
- *solar engineering*
- *coal production*
- *solar water heating*
- *solar hot water system*
- *energy use*
- *air infiltration*
- *sound energy*
- *scientific challenges*
- *primary energy source*
- *residential-commercial sector*
- *watt-hours*
- *wind energy conversion systems*
- *nuclear science and engineering*
- *energia elétrica - qualidade*
- *central and eastern europe*
- *impact energy*
- *greenhouse effect*
- *solar thermal systems*
- *solar thermal*
- *emissions control*
- *building stock*
- *building energy efficiency*
- *emission scenarios*
- *policy scenarios*
- *cell energy*
- *water research*
- *coal combustion*
- *environmentalists*
- *clean coal technologies*
- *energy saves*
- *central asia*
- *external energy*
- *energy storage*
- *conservation of energy*
- *gas plants*
- *biomass potential*
- *sustainability*
- *bioenergy trade*
- *energy exchange*
- *regional energy planning*
- *conservation policies*

A.4 Multi-source term extraction

Terms from Compendex, Inspec, Scirus, back-pointing generation with *renewable energy*, and back-pointing generation with *sustainable energy*:

- *1.27e15 j*
- *450*
- *[jel] q48*
- *aboriginal communities*
- *accelerated depreciation plan*
- *accelerator*
- *access key*
- *active power*
- *actuators*
- *ad 2020-2021*
- *advisory committee*
- *aerodynamic*
- *aerodynamics*
- *after sales service*
- *aggregated wind turbines*
- *agricultural sector*
- *air infiltration*
- *alarm limit*
- *alloy surfaces*
- *alternative energy*
- *alternative energy source*
- *alternative fuel*
- *alternative fuels*
- *american council*
- *amorphous silicon solar cells*
- *annual growth*
- *applied energy*
- *aquifer*
- *aquifers*
- *argentina*
- *articles*
- *asean economic sector*
- *ashrae*
- *atmospheric carbon dioxide*
- *atmospheric sciences*
- *australia*
- *australian cooperative research center for renewable energy*
- *australian cooperative research centre for renewable energy research program*
- *australian renewable energy sector*
- *auxiliary service*
- *awareness generation*
- *bachelor degree program*
- *balanced energy-mix*
- *baltic states*
- *bangladesh*
- *barrels per day*
- *base load*
- *batch reactors*
- *battery consortium*
- *belgium*
- *beryllium disease*
- *best practices*
- *bibliometric measures*
- *binary cycle*
- *bio energies*
- *bio energy*
- *bio-energy systems*
- *biodiesel*
- *bioenergy*
- *bioenergy trade*
- *biofuel*
- *biofuel energy*
- *biofuels*
- *biogas*
- *biogas electricity conversion*
- *biogas plants*
- *biogas system*
- *biological attack*
- *biomass*
- *biomass briquettes*
- *biomass combustion*
- *biomass conversion*
- *biomass energy*
- *biomass energy sources*
- *biomass fuels*
- *biomass gasification*
- *biomass potential*
- *biomass resources*
- *bioresource technology*
- *bioresources*
- *blades*
- *book reviews*
- *borehole*
- *breeding populations*
- *broad policy guidelines*
- *broad-winged hawk*
- *building designers*
- *building energy*
- *building energy efficiency*
- *building industry*
- *building regulations*
- *building stock*
- *buildings*
- *built environment*
- *built environments*
- *bureau of energy of ministry of economical affairs*
- *business plans*
- *business reasons*
- *calculation tools*
- *capital cost*
- *carbon cap-and-trade*
- *carbon dioxide discharges*
- *carbon dioxide emissions*
- *carbon emissions*
- *carbon sequestration*
- *cascade of hydro power plants*
- *case studies*
- *cell energy*
- *cell technology*
- *central and eastern europe*
- *central asia*
- *central australia*
- *centre of appropriate technology*
- *challenge 2008*
- *chemical safety*
- *china*
- *china-india cooperation*
- *chronic beryllium disease*
- *clean coal*
- *clean coal technologies*
- *clean development mechanism (cdm)*
- *clean fuels*
- *climate control*
- *climate policy*
- *climatology*
- *co-operative*
- *co-ordination*

- *co2*
- *co2 emission control policy*
- *coal combustion*
- *coal liquefaction*
- *coal production*
- *coal research*
- *coal-fired power plants*
- *coherent renewable energy policy*
- *collaborative initiatives*
- *combined heat and power*
- *commercial energy scene*
- *commercial risks*
- *commercial sector*
- *commercialisation*
- *commercialization*
- *community*
- *community attitudes*
- *comparative analysis*
- *competing technologies*
- *competitive electricity supply market*
- *competitive power markets*
- *computational science*
- *conservation of energy*
- *conservation policies*
- *construction materials*
- *consumer behaviour*
- *consumer products*
- *contact dr*
- *continuous energy supply*
- *conventional energy sources*
- *conventional generation cost*
- *conventional power*
- *conversion systems*
- *cooking*
- *cooling*
- *correlation*
- *cost reduction*
- *cost/efficiency ratio minimisation*
- *course structure*
- *create jobs*
- *critical loads*
- *cross-flow*
- *crude oil supplies reduction*
- *customer choice programs*
- *data center*
- *de janeiro*
- *decarbonised world*
- *decentralised development*
- *decentralized power systems penetration*
- *decision-making*
- *decision-support system*
- *deep drilling*
- *delphi study*
- *deregulation*
- *desalinated water*
- *desertification problem*
- *design for sustainability*
- *developed countries*
- *developing countries*
- *developing country*
- *development mechanism*
- *development processes*
- *diesel*
- *diesel generators*
- *diffusion of renewables*
- *digester gas*
- *dinosaurs*
- *direct use*
- *distance learning*
- *distributed applications*
- *distributed energy*
- *distributed energy generation*
- *distributed generation*
- *distributed power*
- *distribution networks*
- *distribution of electricity*
- *district heating*
- *doe program*
- *domestic buildings*
- *domestic cooking*
- *domestic energy*
- *domestic production*
- *domestic renewable energy generators*
- *domestic technology manufacturing*
- *donate*
- *donor funding*
- *drilled*
- *drilling operations*
- *drudgery reduction*
- *dry steam*
- *dual fuel engine*
- *dynamic life cycle assessment*
- *dynamic testing*
- *earth climate change*
- *earth science*
- *ecodesign*
- *ecologically optimized transport systems*
- *ecologically sustainable development*
- *economic aspects*
- *economic co-operation*
- *economic comparisons*
- *economic costs*
- *economic developments*
- *economic incentive*
- *economic incentives*
- *economic information*
- *economic performance*
- *economic topic*
- *economical potential*
- *editorials*
- *education and trainings*
- *education programs*
- *educational materials*
- *efficiency*
- *efficiency measures*
- *efficient appliances*
- *efficient lighting*
- *eirev*
- *electric demands*
- *electric energy*
- *electric grid*
- *electric loads*
- *electric power policy frameworks*
- *electric power systems*
- *electric power systems research*
- *electric range*
- *electric vehicle*
- *electric vehicles*
- *electrical and computer engineering (ece)*
- *electrical engineering*
- *electrical power facility (epf)*
- *electrical power test*
- *electricity*
- *electricity consumption*
- *electricity demand*
- *electricity generation*
- *electricity generator*
- *electricity grid*
- *electricity industry*
- *electricity industry restructuring*
- *electricity market reforms*
- *electricity markets*

- *electricity price*
- *electricity prices*
- *electricity production*
- *electricity sales*
- *electricity sector*
- *electricity suppliers*
- *electricity supply*
- *electricity supply industry*
- *electricity use*
- *electrification*
- *electrodialysis*
- *electrolyzer*
- *electroplating*
- *elsevier (co)*
- *emerging renewable energy markets*
- *emission control policy*
- *emission scenarios*
- *emissions (asphalt)*
- *emissions characteristics*
- *emissions control*
- *emissions reduction*
- *emissions trading*
- *employment factors*
- *employment growth*
- *employment opportunities*
- *end-use efficiency*
- *energia elétrica - qualidade*
- *energy*
- *energy and buildings*
- *energy applications*
- *energy assessment methods*
- *energy audit*
- *energy capacity*
- *energy carrier*
- *energy certificate*
- *energy conservation*
- *energy conservation efforts*
- *energy conservation measures*
- *energy consumption*
- *energy conversion and management*
- *energy conversion efficiency*
- *energy conversion systems*
- *energy converters*
- *energy crisis*
- *energy crops*
- *energy demand*
- *energy demand increase*
- *energy demand variations*
- *energy demands*
- *energy density utilization*
- *energy development*
- *energy diversification*
- *energy economics*
- *energy economics and policy*
- *energy efficiency*
- *energy efficient*
- *energy evaluation*
- *energy exchange*
- *energy extraction*
- *energy farming*
- *energy generation*
- *energy impacts*
- *energy industry*
- *energy intensities*
- *energy intensive economy*
- *energy management*
- *energy market*
- *energy mix*
- *energy model*
- *energy output*
- *energy performance regulations*
- *energy planning*
- *energy policies*
- *energy policy*
- *energy policy act*
- *energy politics*
- *energy potential*
- *energy prices*
- *energy production*
- *energy products*
- *energy requirement*
- *energy research*
- *energy resource*
- *energy resources*
- *energy saves*
- *energy savings*
- *energy sector*
- *energy security*
- *energy services*
- *energy shortages*
- *energy situation*
- *energy sources*
- *energy storage*
- *energy studies*
- *energy supplies*
- *energy supply*
- *energy supply security considerations*
- *energy surplus*
- *energy system analysis*
- *energy systems*
- *energy systems engineering*
- *energy technologies*
- *energy technology*
- *energy unit*
- *energy usage*
- *energy use*
- *energy utilities*
- *energy utilization*
- *energy-efficient lighting*
- *energy-generating resources*
- *energy-price signals reformation*
- *energy-saving technical progress*
- *engineering program*
- *enhanced greenhouse effect*
- *enthalpy*
- *environment issues*
- *environment programme*
- *environmental benefits*
- *environmental conservation*
- *environmental contamination*
- *environmental cost*
- *environmental costs*
- *environmental education*
- *environmental engineering*
- *environmental friendly sustainable development*
- *environmental groups*
- *environmental impact assessment*
- *environmental impacts*
- *environmental interest*
- *environmental issues*
- *environmental performance*
- *environmental policies*
- *environmental policy*
- *environmental pollution*
- *environmental pollutions*
- *environmental problems*
- *environmental protection agency*
- *environmental superiority*
- *environmental systems*
- *environmental technology*
- *environmentalism*
- *environmentalists*

- *environmentally friendly alternatives*
- *environmentally friendly energy supply*
- *estonia*
- *ethanol*
- *eu energy performance of buildings directive*
- *european islands*
- *european renewable energy council (erec)*
- *european union*
- *executive director*
- *exergy*
- *exergy analysis*
- *exergy efficiency*
- *experience curves*
- *external energy*
- *fahrenheit*
- *failure analysis*
- *fall in price*
- *farm operation*
- *fatigue life*
- *fault detection*
- *federal agencies*
- *federal commitment*
- *final energy*
- *finance pre-feasibility studies*
- *financial assistance*
- *financial incentives*
- *financial viability*
- *financing*
- *financing costs*
- *financing processes*
- *financing risk premiums reduction*
- *financing support*
- *finite energy resources*
- *fiscal year*
- *fission*
- *flexible photo-voltaic solar cell*
- *food technology*
- *form of energy*
- *fort collins*
- *fossil fuel*
- *fossil fuel combustion*
- *fossil fuel economy*
- *fossil fuel electricity generation*
- *fossil fuel politics*
- *fossil fuels*
- *fractures*
- *framework programme*
- *franchise tax*
- *freedom of information*
- *fuel cell*
- *fuel cell systems*
- *fuel cells*
- *fuel charge*
- *fuel cycle*
- *fuel economy*
- *fuel efficiency*
- *fuel for thought strategies*
- *fuel poverty*
- *fuel power*
- *fuel savings*
- *fuel standards*
- *fuel storage*
- *fuel transportation*
- *funding mechanism*
- *fusion energy*
- *future energy*
- *future energy supply networks*
- *future generations*
- *galapagos archipelago*
- *gas plants*
- *gas reserves*
- *gas supply*
- *gas-fired*
- *gas-tight*
- *gasoline*
- *gathering information*
- *gearbox*
- *general electric*
- *general public*
- *generation cost*
- *generation of electricity*
- *generation systems*
- *genoa*
- *geoscience*
- *geothermal*
- *geothermal activity*
- *geothermal area*
- *geothermal areas*
- *geothermal development*
- *geothermal drilling*
- *geothermal energy*
- *geothermal exploration*
- *geothermal field*
- *geothermal fields*
- *geothermal fluids*
- *geothermal gradient*
- *geothermal heat pumps*
- *geothermal park*
- *geothermal plant*
- *geothermal power*
- *geothermal power plant*
- *geothermal power plants*
- *geothermal reservoir*
- *geothermal reservoirs*
- *geothermal resources*
- *geothermal resources -- environmental aspects -- oregon -- prineville region*
- *geothermal sources*
- *geothermal steam*
- *geothermal system*
- *geothermal systems*
- *geothermal water*
- *geothermal waters*
- *geothermal wells*
- *german renewable energy source act*
- *germany*
- *geyser*
- *gibraltar strait*
- *global climate change*
- *global energy demand*
- *global issues*
- *global markets*
- *global population*
- *global renewable energy industry*
- *global renewable energy movement*
- *global security*
- *global warming*
- *golden eagles*
- *government*
- *government documents*
- *government incentives*
- *government mandates*
- *government of india*
- *government policies*
- *government policy decisions*
- *government subsidies*
- *government subsidy*
- *government support*
- *great basin*
- *great plains*
- *greece*
- *green certificates*
- *green electricity*
- *green pricing*
- *greenhouse climate control*
- *greenhouse effect*

- *greenhouse gas*
- *greenhouse gas abatement*
- *greenhouse gas emissions*
- *greenhouse gas reduction*
- *greenhouse gases*
- *greenhouse heating*
- *greenhouse issues*
- *grid systems*
- *grid-connected systems*
- *ground loop*
- *grounding*
- *growing economy*
- *growth trends*
- *harmful emissions*
- *harnessing*
- *heat energy*
- *heat exchangers*
- *heat pump*
- *heat pump system*
- *heat pumps*
- *heat supply system*
- *heated*
- *heating*
- *heating system*
- *heating systems*
- *heating water*
- *high activity subjects*
- *high cost*
- *high-altitude*
- *high-pressure*
- *higher educational institutions*
- *hindu kush himalayan region*
- *home energy*
- *horizontal axis*
- *hot spring*
- *hot springs*
- *hot water systems*
- *household energy situation*
- *human genome*
- *human genome project*
- *human health*
- *hybrid development*
- *hybrid diesel systems*
- *hybrid power systems*
- *hybrid renewable energy systems*
- *hybrid solar-wind system*
- *hydraulic*
- *hydro power*
- *hydro-electric power*
- *hydroelectric*
- *hydroelectric dams*
- *hydroelectric energy*
- *hydroelectric power*
- *hydroelectricity*
- *hydrogen*
- *hydrogen energy*
- *hydrogen fuel*
- *hydrogen infrastructure*
- *hydrogen systems*
- *hydrogen vehicles*
- *hydrology*
- *hydropower*
- *hydrothermal*
- *hydrothermal systems*
- *ignalina nuclear power plant*
- *impact energy*
- *impact on the environment*
- *incentive measures*
- *incentive mechanism*
- *incubator*
- *india*
- *indian national programme*
- *indian renewable energy development agency*
- *indian renewable energy development agency limited*
- *industrial ecology*
- *industrial energy*
- *industrial sector*
- *industrial technologies*
- *information service*
- *innovation systems*
- *innovative business models*
- *installed capacity*
- *installed power*
- *institutional barriers*
- *institutional structures*
- *institutional support*
- *insulation*
- *integrated assessment*
- *integrated gasification*
- *integrated resource*
- *integration system*
- *intelligent software*
- *intergovernmental panel*
- *intergovernmental panel on climate change*
- *intermittent*
- *intermittent renewable energy penetration*
- *international agreements*
- *international bank for reconstruction and development*
- *international conferences*
- *international cooperation*
- *international energy agency*
- *international initiatives*
- *international law*
- *international network*
- *international renewable energy development*
- *internationally focused*
- *internet*
- *intrinsic technology problem*
- *investment*
- *investment decisions*
- *investment deduction*
- *investment incentives*
- *investments financing*
- *investor*
- *ipcc*
- *irrigation water pumping*
- *islands*
- *issue date*
- *japan*
- *job losses*
- *joint ventures*
- *kilowatts*
- *kirklees councils*
- *kizildere*
- *land management*
- *landfill gas*
- *large scale dissemination*
- *latvia*
- *lca*
- *lead-acid batteries*
- *lead-acid battery*
- *legal barriers*
- *legal environment*
- *legislation*
- *legislative aspects*
- *letters*
- *levelized energy costs*
- *licensing procedure*
- *life cycle assessment*
- *life-cycle*
- *light energy*
- *lighting*
- *lighting end-uses*
- *linear generator*
- *liquefaction*
- *liquid column*

- *liquid dampers*
- *lithuanian energy sector*
- *little research*
- *livermore national laboratory*
- *load data*
- *load factor*
- *load management*
- *load powering*
- *local authorities*
- *local employments*
- *local government*
- *local processing*
- *local residents*
- *local wind*
- *long-term contracts*
- *long-term stability*
- *low activity subjects*
- *low carbon*
- *low energy density*
- *low head*
- *low wind speed turbine*
- *low-income families*
- *low-interest loans*
- *maintenance*
- *malaysia*
- *maldives*
- *maps*
- *marginalized mountain population*
- *marine renewable energy*
- *market adoptions*
- *market barriers*
- *market participants*
- *market restructuring*
- *market transformation*
- *materials handling*
- *materials technology*
- *mechanical engineering*
- *mechanical power*
- *mechatronics*
- *media relations*
- *megawatts*
- *metal recovery*
- *methane*
- *methane generation*
- *methane yield*
- *mexico*
- *micro hydropower*
- *microforms*
- *microhydro*
- *miles per hour*
- *milieu*
- *milieukunde*
- *millennium development goals (mdg)*
- *ministry of economic affairs of taiwan*
- *ministry of non-conventional energy sources*
- *ministry of nonconventional energy sources*
- *mitigation*
- *modelling study*
- *multidisciplinary programs*
- *multilevel search engines*
- *municipal utilities*
- *murdoch university*
- *nameplate*
- *nanotechnology*
- *national economy*
- *national energy planning*
- *national parks*
- *national power installation capacity*
- *national security*
- *natural ecosystems*
- *natural gas consumption*
- *natural gas supplies reduction*
- *negative effects*
- *negative secondary impacts*
- *nepal*
- *new zealand*
- *non fossil fuel obligation*
- *non-governmental organizations*
- *non-renewable*
- *non-renewable energy*
- *nongovernmental organizations*
- *nonpooled electricity trading*
- *northern australia*
- *northern queensland*
- *northern south australia*
- *northern territory*
- *notes*
- *novel design*
- *nuclear data*
- *nuclear research*
- *nuclear safety*
- *nuclear science*
- *nuclear science and engineering*
- *nuclear technology*
- *nuclear weapons*
- *ocean breezes*
- *ocean energy*
- *ocean wave energy*
- *office of basic energy sciences*
- *office of energy research*
- *office of science*
- *offshore environment*
- *offshore wind turbine*
- *oil*
- *oil crisis*
- *oil prices*
- *oil reserves*
- *oil supply*
- *onshore*
- *opec effect*
- *open loop control system*
- *optimal placement*
- *optimal renewable energy mathematical (orem) model*
- *optimal renewable energy mathematical model*
- *optimal renewable energy model*
- *optimization*
- *optimization model*
- *optimization models*
- *option to purchase*
- *organisation for economic co-operation and development*
- *organizational aspects*
- *organizational development*
- *outlays*
- *output-based allocation*
- *overview of fuel cell*
- *ownership system*
- *pakistan*
- *participation*
- *pastoral properties*
- *peak energies*
- *peak powers*
- *pelton wheel*
- *performance indicators*
- *performance prediction*
- *petroleum crude oil consumption*
- *petroleum gas*
- *petroleum reserves*
- *petroleum resources*
- *photovoltaic*

- *photovoltaic array*
- *photovoltaic cells*
- *photovoltaic energy*
- *photovoltaic generation system*
- *photovoltaic system*
- *photovoltaic systems*
- *photovoltaic technology*
- *photovoltaics*
- *physical education*
- *pitch angle*
- *planning*
- *planning phases*
- *plasma physics*
- *plate tectonics*
- *plutonium*
- *policies*
- *policies and measures to enhance use of renewable energy sources*
- *policies formulation*
- *policy design impacts*
- *policy incentives*
- *policy makers*
- *policy making*
- *policy objectives*
- *policy options*
- *policy reform*
- *policy scenarios*
- *pollutant emissions*
- *polluting*
- *postgraduate program*
- *postgraduate university courses*
- *potential resources*
- *poultry litter*
- *power coefficient*
- *power data*
- *power development*
- *power electronics*
- *power generation*
- *power grid interconnection*
- *power market*
- *power plant financing process*
- *power produced*
- *power production*
- *power quality*
- *power resource*
- *power sector*
- *power sectors*
- *power sources*
- *power stations*
- *power supply*
- *power system control*
- *power system design*
- *power system operation*
- *power system planning*
- *power systems*
- *power systems engineering*
- *power utility*
- *present situation*
- *pressure decline*
- *price drops*
- *price mechanism*
- *price of energy*
- *primary energy source*
- *primary factors*
- *private sector participation*
- *producer gas*
- *product design*
- *production of hydrogen*
- *production systems*
- *production tax*
- *progress in energy and combustion science*
- *progressive utilization*
- *project datums*
- *project development*
- *project leaders*
- *projects*
- *promotion of renewable energy sources*
- *promotion strategies*
- *propulsion*
- *public acceptance*
- *public affairs*
- *public benefits*
- *public lands -- oregon -- prineville region -- management*
- *public participation*
- *public perceptions*
- *publications*
- *pump system*
- *pump systems*
- *pumping*
- *pv*
- *pv power systems*
- *pv-compact*
- *qualitative data*
- *quality control*
- *quality of life*
- *questionnaire surveys*
- *quota-based systems*
- *r&d*
- *r&d support*
- *rankine cycle*
- *rated capacity*
- *rated power*
- *rd and d*
- *reactive power*
- *record search*
- *recursive linear planning model*
- *recursive linear program*
- *reducing greenhouse gas emissions*
- *reduction of carbon dioxide*
- *refurbishment projects*
- *regional development*
- *regional energy planning*
- *regional power market competition*
- *regulated rural energy concessions*
- *regulating*
- *regulatory law framework*
- *reliability*
- *reliability constraints*
- *reliable approach*
- *remote area power supplies*
- *remote areas*
- *remote power generation*
- *renewable electricity generation*
- *renewable energies*
- *renewable energy*
- *renewable energy (re)*
- *renewable energy (re) technologies*
- *renewable energy activities*
- *renewable energy assessment*
- *renewable energy capacity*
- *renewable energy centres*
- *renewable energy commercialization*
- *renewable energy company limited*
- *renewable energy costs*
- *renewable energy curriculum*
- *renewable energy development*
- *renewable energy development bill*
- *renewable energy development plan*

- *renewable energy development status*
- *renewable energy development strategies*
- *renewable energy developments*
- *renewable energy distribution*
- *renewable energy education*
- *renewable energy engineering curriculum*
- *renewable energy engineering system*
- *renewable energy equipment procurement*
- *renewable energy financing*
- *renewable energy generating equipments*
- *renewable energy generation*
- *renewable energy generation cost*
- *renewable energy incentives*
- *renewable energy industry*
- *renewable energy investment*
- *renewable energy investments*
- *renewable energy law*
- *renewable energy literature*
- *renewable energy market*
- *renewable energy markets*
- *renewable energy materials*
- *renewable energy penetration*
- *renewable energy plan*
- *renewable energy plan 2012*
- *renewable energy planning*
- *renewable energy policies*
- *renewable energy policy*
- *renewable energy policy harmonisation*
- *renewable energy policy options*
- *renewable energy potential*
- *renewable energy product purchase*
- *renewable energy program*
- *renewable energy programmes*
- *renewable energy projects*
- *renewable energy promotion law*
- *renewable energy promotion strategies*
- *renewable energy quota system*
- *renewable energy r&d*
- *renewable energy research*
- *renewable energy resources*
- *renewable energy scenario*
- *renewable energy sector*
- *renewable energy sources*
- *renewable energy sources (res)*
- *renewable energy standards*
- *renewable energy statistics*
- *renewable energy subsidy policy*
- *renewable energy supply (res)*
- *renewable energy systems*
- *renewable energy systems allocation*
- *renewable energy systems cost*
- *renewable energy systems reliability*
- *renewable energy technologies*
- *renewable energy technology*
- *renewable energy technology (ret)*
- *renewable energy technology research*
- *renewable energy trading experience*
- *renewable energy usage*
- *renewable energy utilisation*
- *renewable energy utilization*
- *renewable facilites*
- *renewable portfolio*
- *renewable portfolio standards*
- *renewable resource*
- *renewable sources*
- *renewable system*
- *renewables*
- *renewislands*
- *replenish*
- *republic of croatia*
- *research and development*
- *research and developments*
- *research efforts*
- *research output*
- *research priorities*
- *reservoir simulation*
- *residential building*
- *residential buildings*
- *residential-commercial sector*
- *resource assessment*
- *resource potential*
- *ress*
- *restructured electricity market*
- *revenue certainty*
- *reviews*
- *rhyolite*
- *ring-fenced markets*
- *rinse water*
- *roaring forties*
- *rock formations*
- *role of hydrogen*
- *role of particles*
- *role of technology*
- *rotor blades*
- *rotor diameter*
- *rotor speed*
- *rural areas*
- *rural communities*
- *rural economy*
- *rural electrification*
- *rural electrification in senegal*
- *rural energy*
- *rural energy enterprises*
- *rural energy services*
- *rural livelihoods*
- *salton sea*
- *salton trough*
- *sandstone*
- *satellite-based sensors*
- *scale-up*
- *scaling up*
- *scheikunde*
- *school environment*
- *scientific challenges*
- *secretary of state*
- *semi-empirical models*
- *semi-structured interviews*
- *senegal*

- *senior vice president*
- *sensitivity analysis*
- *sequencing batch*
- *sequencing batch reactors*
- *service provider*
- *short segment*
- *significant wave height*
- *siting*
- *small communities*
- *small island developing states*
- *small wind turbines*
- *small-scale renewable energy technologies*
- *social acceptance*
- *social acceptance variation*
- *social issues*
- *social studies*
- *socio economic optimal renewable energy model*
- *soil gas*
- *solar cell*
- *solar city*
- *solar collectors*
- *solar concentrator*
- *solar cooker*
- *solar electricity*
- *solar energy*
- *solar energy applications*
- *solar energy conversion*
- *solar energy materials and solar cells*
- *solar energy systems*
- *solar energy utilization*
- *solar engineering*
- *solar heat pump combination*
- *solar heating*
- *solar heating system*
- *solar hot water system*
- *solar house*
- *solar hydrogen*
- *solar insolation*
- *solar lanterns*
- *solar photovoltaic*
- *solar photovoltaics*
- *solar power*
- *solar power plant*
- *solar pv*
- *solar pv utilization*
- *solar regime*
- *solar systems*
- *solar thermal*
- *solar thermal energy*
- *solar thermal systems*
- *solar water*
- *solar water heaters*
- *solar water heating*
- *solar water heating systems*
- *solar wind*
- *sound energy*
- *sources of energy*
- *space heating*
- *space technologies*
- *spain*
- *spatial planning*
- *spring water*
- *stand-alone*
- *stand-alone power systems*
- *standardization*
- *state energy*
- *state government*
- *state government agencies*
- *state of michigan*
- *statute for upgrading industries*
- *steam power*
- *steam production*
- *stingray tidal current generator*
- *stingray tidal stream generator*
- *storage medium*
- *storage technologies*
- *storage technology*
- *storm damage*
- *strait*
- *stream power*
- *strong winds*
- *subcontractor*
- *subsidy measures*
- *substitute cost*
- *subsurface*
- *supercomputing*
- *supply shortages*
- *support mechanisms*
- *sustainability*
- *sustainable buildings*
- *sustainable development*
- *sustainable energy*
- *sustainable energy source*
- *sustainable energy sources*
- *sustainable future*
- *sustainable growth*
- *sustainable use*
- *sustained renewable energy deployment*
- *swine manure*
- *system voltage*
- *systems research*
- *taiwan*
- *tariff*
- *tasmania*
- *tax credit*
- *tax incentives*
- *tax revenue*
- *tax revenue method*
- *taxpayer*
- *technical potential*
- *technological aspects*
- *technological development*
- *technological learning*
- *technological research*
- *technology choice*
- *technology division*
- *technology push*
- *technology transfer*
- *terrorism*
- *tertiary education*
- *tertiary renewable energy education*
- *tertiary trained engineers*
- *tertiary trained policy makers*
- *tertiary trained scientists*
- *test reactor*
- *testing phase*
- *thermal power*
- *thermal spring*
- *thermal springs*
- *thermal systems*
- *thermal water*
- *thermal waters*
- *thermochemical conversion*
- *thermonuclear power*
- *tidal barrage*
- *tidal channel*
- *tidal current*
- *tidal elevation*
- *tidal energy*
- *tidal power*
- *tidal power station*
- *tidal stream*
- *tidal wave*
- *time 2020 year to 2021 year*
- *time-dependent*
- *time-dependent systems*

- *tlcd*
- *top-down*
- *total energies*
- *total energy demand*
- *tourist ventures*
- *transient modelling*
- *transmission networks*
- *transport sector*
- *transportation*
- *transportation fuels*
- *transportation sector*
- *triple point*
- *turbine*
- *turbine blades*
- *turbine generator*
- *turbine rotor*
- *turbines*
- *turbulent*
- *turkey*
- *u.s. energy demand*
- *uk*
- *uk nffo*
- *undergraduate education*
- *undergraduate university courses*
- *underwater power generation*
- *unemployment rates*
- *uniform green power market mode*
- *united nations development program*
- *united nations environment programme*
- *united nations industrial development organization*
- *uranium*
- *uranium mining*
- *urban areas*
- *urban settings*
- *usa*
- *use of hydrogen*
- *vibration control*
- *vice president*
- *virgin islands*
- *volcanic zone*
- *volcano*
- *wairakei*
- *wairakei field*
- *wake effects*
- *warm springs*
- *washington*
- *waste energy resources*
- *waste heat*
- *water demands*
- *water heaters*
- *water heating*
- *water pumping*
- *water research*
- *water turbine*
- *watt-hours*
- *wave direction*
- *wave energy*
- *wave energy conversion*
- *wave loading*
- *wave power*
- *web-based renewable energy courses*
- *welding*
- *west texas*
- *western australia*
- *western flank*
- *wind*
- *wind atlas*
- *wind energy*
- *wind energy conversion*
- *wind energy conversion systems*
- *wind energy converter*
- *wind energy potential*
- *wind energy utilization*
- *wind events*
- *wind excitation*
- *wind facilities*
- *wind farm*
- *wind farms*
- *wind generation*
- *wind loading*
- *wind plants*
- *wind power*
- *wind power density*
- *wind power development*
- *wind power generation*
- *wind power plant*
- *wind resource*
- *wind resource assessment*
- *wind speed data*
- *wind system*
- *wind systems*
- *wind turbine*
- *wind turbines*
- *wind velocity*
- *windfarm market*
- *wood waste*
- *working fluid*
- *workshop in a box program*
- *world bank*
- *world leaders*
- *world wide web*

A.5 Eigenvector centrality relevance testing

Eigenvector centrality rankings, 1,000-term network:

- *renewable energy*
- *renewable energy sources*
- *renewable sources*
- *renewable energy technologies*
- *energy technologies*
- *renewable energy systems*
- *renewable energy resources*
- *renewable energies*
- *renewables*
- *wind power*
- *geothermal energy*
- *energy resource*
- *wind energy*
- *energy development*
- *energy resources*
- *rural energy*
- *energy systems*
- *wind power development*
- *alternative energy*
- *power generation*
- *generation of electricity*
- *solar electricity*
- *energy planning*
- *energy technology*
- *sources of energy*
- *energy situation*
- *wind energy potential*
- *sustainable energy*
- *microhydro*
- *electricity grid*
- *energy policy*
- *renewable system*
- *future energy*
- *biomass energy*
- *rural electrification*
- *energy*
- *solar power*
- *installed power*
- *energy efficiency*
- *geothermal development*
- *electric energy*
- *geothermal sources*
- *energy sector*
- *geothermal power*
- *geothermal*
- *hydro power*
- *wind resource*
- *wind turbines*
- *electricity generation*
- *energy industry*
- *state energy*
- *electricity production*
- *energy impacts*
- *small wind turbines*
- *power sector*
- *wind farm*
- *energy economics*
- *electricity supply*
- *energy supplies*
- *electricity*
- *wind*
- *distributed generation*
- *wind systems*
- *solar photovoltaics*
- *energy services*
- *industrial energy*
- *geothermal resources*
- *wind farms*
- *electricity sector*
- *energy system analysis*
- *megawatts*
- *energy research*
- *photovoltaic systems*
- *wind turbine*
- *solar energy applications*
- *wind generation*
- *end-use efficiency*
- *turbine*
- *installed capacity*
- *biomass resources*
- *solar energy utilization*
- *turbines*
- *geothermal plant*
- *hydropower*
- *remote power generation*
- *energy capacity*
- *energy crisis*
- *electric grid*
- *wave energy*
- *energy management*
- *solar energy systems*
- *energy potential*
- *domestic energy*
- *energy products*
- *experience curves*
- *electrification*
- *tidal energy*
- *tidal power*
- *geothermal power plant*
- *building energy*
- *power development*
- *emissions reduction*
- *development mechanism*
- *district heating*
- *exergy efficiency*
- *home energy*
- *amount of energy*
- *solar power plant*
- *electricity supply industry*
- *heat energy*
- *geothermal power plants*
- *electric utilities*
- *exergy*
- *energy conservation*
- *hydroelectric*
- *solar systems*
- *emissions trading*
- *clean coal*
- *bioenergy*
- *reducing greenhouse gas emissions*
- *improved energy*
- *wind system*
- *energy savings*
- *renewable resource*
- *electricity use*
- *kilowatts*
- *geothermal steam*
- *energy output*
- *pv systems*
- *geothermal fields*
- *low wind speed turbine*
- *distribution of electricity*
- *hydrogen fuel*
- *energy utilization*
- *non-renewable*
- *energy usage*
- *wind energy conversion*
- *solar house*
- *electricity industry*
- *geothermal systems*
- *power sources*
- *photovoltaic power*

- *production tax*
- *greenhouse gas emissions*
- *emissions reductions*
- *energy-efficient lighting*
- *solar water*
- *carbon emissions*
- *photovoltaics*
- *form of energy*
- *cogeneration*
- *fuel cell*
- *efficient appliances*
- *geothermal water*
- *greenhouse gas*
- *biomass conversion*
- *geothermal waters*
- *electricity demand*
- *climate policy*
- *storage technologies*
- *energy economics and policy*
- *climate change policy*
- *thermal waters*
- *energy efficient*
- *co2 emissions*
- *transportation sector*
- *water heating*
- *rated power*
- *energy performance*
- *electric power systems*
- *solar hydrogen*
- *conversion systems*
- *energy-saving technical progress*
- *solar energy conversion*
- *solar power generation*
- *energy prices*
- *technological learning*
- *emission reductions*
- *geothermal fluids*
- *electric system*
- *sustainable development*
- *solar heating*
- *biomass*
- *wind energy converter*
- *heating systems*
- *tidal stream*
- *rotor diameter*
- *electricity consumption*
- *municipal utilities*
- *energy crops*
- *fuel cells*
- *alternative fuels*
- *energy audit*
- *buildings*
- *geothermal reservoirs*
- *photovoltaic system*
- *carbon intensity*
- *fuel efficiency*
- *turbine generator*
- *market barriers*
- *electricity prices*
- *international emissions*
- *water heaters*
- *wind power plant*
- *non-renewable resources*
- *thermochemical conversion*
- *water pumping*
- *harnessing*
- *energy intensity*
- *geothermal reservoir*
- *wind engineering*
- *excess power*
- *commercial buildings*
- *geothermal system*
- *efficiency*
- *generating station*
- *power produced*
- *power demand*
- *electricity markets*
- *power production*
- *space heating*
- *deregulation*
- *policy instruments*
- *kyoto protocol*
- *geothermal wells*
- *wairakei field*
- *resource potential*
- *landfill gas*
- *wairakei*
- *geothermal area*
- *greenhouse heating*
- *geothermal exploration*
- *emission trading*
- *generating units*
- *wind speed data*
- *geothermal areas*
- *transportation fuels*
- *economical*
- *thermal power*
- *resource assessment*
- *carbon dioxide emissions*
- *hydrogen*
- *solar collectors*
- *sustainable future*
- *small island developing states*
- *power quality*
- *photovoltaic*
- *emission reduction*
- *thermal systems*
- *electric generator*
- *integrated gasification*
- *geothermal field*
- *solar heat*
- *integrated resource*
- *electricity distribution*
- *energy crop*
- *residential buildings*
- *solar generator*
- *offshore environment*
- *greenhouse gas abatement*
- *power coefficient*
- *nox emissions*
- *irrigation water pumping*
- *solar cell*
- *heat pumps*
- *horizontal axis*
- *hydroelectric dams*
- *geothermal activity*
- *gas-fired*
- *harmful emissions*
- *gas supply*
- *abatement*
- *low carbon*
- *combustion technology*
- *environmental benefits*
- *geothermal gradient*
- *emissions permits*
- *steam power*
- *oil supply*
- *transport sector*
- *generator system*
- *thermal wind*
- *heating system*
- *heat pump*
- *geothermal gradients*
- *turbine blades*
- *turbine noise*
- *base load*
- *non-polluting*
- *light energy*
- *outer continental shelf*
- *climate control*
- *non-renewable resource*
- *coal-fired plants*
- *power rating*
- *office of basic energy sciences*
- *nameplate*

- *domestic hot water*
- *onshore*
- *heat pump system*
- *electricity generator*
- *trading program*
- *space cooling*
- *electric power transmission*
- *biomass fuels*
- *solar cooker*
- *turbine rotor*
- *oil crisis*
- *load management*
- *gas turbine*
- *general electric*
- *reactive power*
- *low head*
- *energy recovery*
- *rotor blades*
- *institutional structures*
- *power electronics*
- *doe program*
- *cell technology*
- *pump systems*
- *gasification*
- *impulse turbine*
- *gas industry*
- *bioelectric energy sources*
- *fuel poverty*
- *blades*
- *nonrenewable resources*
- *turbine blade*
- *rotor blade*
- *ventilation air*
- *role of hydrogen*
- *resource and energy economics*
- *wind events*
- *dynamic stall*
- *global climate change*
- *waste heat*
- *variable pitch*
- *coal-fired power plants*
- *siting*
- *capacity factor*
- *short rotation*
- *biomass gasification*
- *life cycle assessment*
- *gas turbines*
- *environmental affairs*
- *power pack*
- *golden eagles*
- *gas reserves*
- *lca*
- *power equipment*
- *ground loop*
- *efficiency measures*
- *battery storage*
- *ashrae*
- *condition monitoring*
- *microbial fuel cell*
- *biogas plants*
- *maldives*
- *pitch angle*
- *farm operation*
- *[jel] q32*
- *industrial technologies*
- *blade root*
- *solar cookers*
- *oil equivalent*
- *business reasons*
- *taupo volcanic zone*
- *international energy agency*
- *endogenous technical change*
- *air emissions*
- *[jel] q3*
- *fuel economy*
- *biogas*
- *extraction policy*
- *portfolio approach*
- *maximum power density*
- *petroleum resources*
- *oil reserves*
- *iron and steel industry*
- *reduction of carbon dioxide*
- *biomass combustion*
- *ecological footprint*
- *solar collector*
- *global carbon*
- *eigenvalue analysis*
- *industrial gas*
- *central heating*
- *petroleum reserves*
- *ocean breezes*
- *roaring forties*
- *industrial gas turbine*
- *carbon sequestration*
- *hvac*
- *mitigation*
- *collector system*
- *extracted resources*
- *electrical network*
- *digester gas*
- *executive director*
- *thermal storage*
- *intergovernmental panel on climate change*
- *power take-off*
- *photovoltaic cell*
- *distribution networks*
- *resource extraction*
- *photovoltaic cells*
- *green taxes*
- *hvac system*
- *siemens*
- *built environment*
- *thermochemical*
- *resource depletion*
- *use efficiency*
- *rotor speed*
- *fatigue loads*
- *dry steam*
- *nuclear industry*
- *flash steam*
- *steam turbines*
- *reliability*
- *oil drilling*
- *mechanical power*
- *environmental costs*
- *bioelectricity*
- *capital cost*
- *hot-humid climate*
- *emissions inventory*
- *water turbine*
- *tidal basin*
- *sustainable welfare*
- *polarization curve*
- *turbine engine*
- *sequestration*
- *wind velocity*
- *oil prices*
- *systems research*
- *performance prediction*
- *wood waste*
- *gas exploration*
- *integrated assessment*
- *steam production*
- *battery backup*
- *environmental sustainability*
- *mineral resources*
- *long-term contracts*
- *mass conservation*
- *energy reserves*
- *storage technology*
- *aerodynamics*
- *gas turbine engine*
- *steam turbine*

- *biofuel*
- *flat terrain*
- *guide vanes*
- *environmental policy*
- *exhaustible resources*
- *oil*
- *environmental groups*
- *environmental economics*
- *hot springs*
- *active power*
- *environmental performance*
- *national wildlife refuge*
- *polluting*
- *jet engine*
- *voltage control*
- *center stage*
- *volcanic zone*
- *oil companies*
- *susceptance*
- *film cooling*
- *germany*
- *environmental impact assessment*
- *power supplies*
- *business climate*
- *liquid fuel*
- *geyser*
- *novel design*
- *increasing returns*
- *nature conservation*
- *diesel*
- *atmospheric emissions*
- *framework convention*
- *innovation policy*
- *market simulation*
- *turbine engines*
- *exhaustible*
- *forest biomass*
- *propulsion*
- *spot market*
- *refocus*
- *environmental systems*
- *environment programme*
- *tc hydraulic engineering ocean engineering*
- *decision support system*
- *ethanol*
- *minnesota*
- *feedstock*
- *thermal water*
- *external cost*
- *brewery*
- *solid oxide fuel cells*
- *pitch diameter*
- *lead-acid batteries*
- *simple cycle*
- *open-circuit voltage*
- *heating oil*
- *gas turbine engines*
- *atmospheric concentrations*
- *greening*
- *united nations environment programme*
- *angle of attack*
- *ethanol production*
- *methane emissions*
- *rocket engines*
- *island press*
- *plant design*
- *desalination*
- *tax breaks*
- *environmental conservation*
- *supplying*
- *steel production*
- *system voltage*
- *loan guarantee*
- *ice accretion*
- *coalbed methane*
- *producer gas*
- *cooling air*
- *mitigation measures*
- *salton trough*
- *agricultural residues*
- *electrolysis*
- *combustors*
- *tariff*
- *internal resistance*
- *conservation reserve program*
- *biodiesel*
- *industrial ecology*
- *changing climate*
- *impacts of climate change*
- *pump system*
- *future generations*
- *hydrogen storage*
- *heat storage*
- *utilization factor*
- *electric motor*
- *building services*
- *stationary sources*
- *original equipment manufacturers*
- *tax incentives*
- *methanol*
- *developing countries*
- *heat exchangers*
- *economic affairs*
- *seasonal demand*
- *alternator*
- *nanotechnology*
- *aerodynamic*
- *environmental education*
- *vehicle fleet*
- *gasoline*
- *scheikunde*
- *planetary gear*
- *specific speed*
- *vice president*
- *sectoral*
- *semi-endogenous growth*
- *wilderness areas*
- *nuclear safety*
- *flowmeter*
- *externalities*
- *crop residues*
- *cathodes*
- *fermentation*
- *tax credit*
- *heat rejection*
- *o emissions*
- *combustion chamber*
- *market mechanisms*
- *sea trials*
- *office of science*
- *cathode*
- *agricultural waste*
- *role of technology*
- *learning rates*
- *turbine flowmeter*
- *optimal placement*
- *national security*
- *gas field*
- *public affairs*
- *radiant*
- *intergenerational equity*
- *hot spring*
- *west texas*
- *fuel charge*
- *paper company*
- *slip rings*
- *reformulated gasoline*
- *economic systems*
- *cost reduction*
- *issue date*
- *emission factors*
- *global markets*
- *territorial*
- *intelligent software*

- *computational science*
- *nitrous oxide emissions*
- *co-ordination*
- *republic of croatia*
- *borehole*
- *ignimbrite*
- *domestic production*
- *bipartisan*
- *impact on the environment*
- *trading systems*
- *warm springs*
- *mechatronics*
- *waste land*
- *cooling system*
- *municipal solid waste*
- *gear box*
- *steel grade*
- *ecological economics*
- *savings plan*
- *usa*
- *environmental law*
- *battery consortium*
- *drilled*
- *rhyolite*
- *government entities*
- *enter into force*
- *organic solar cells*
- *economics association*
- *policy recommendations*
- *environmental justice*
- *exhaust system*
- *mass flow rate*
- *gaseous fuel*
- *gearbox*
- *sensitivity analysis*
- *volcano*
- *methane*
- *atmospheric carbon dioxide levels*
- *insulation*
- *biodiesel production*
- *atmospheric gases*
- *market trends*
- *environmental engineering*
- *balanced growth path*
- *particle erosion*
- *pellet production*
- *gas fields*
- *military aircraft*
- *industrial park*
- *capital outlay*
- *frequency control*
- *intermediate pressure*
- *trade group*
- *land-use*
- *environment agency*
- *nitrous oxide emission*
- *hydrofluorocarbons*
- *inventories*
- *ceramic matrix*
- *high temperature fatigue*
- *american council*
- *sulphur hexafluoride*
- *price increase*
- *logic control*
- *rate of extraction*
- *rechargeable batteries*
- *environmental sciences*
- *ignition system*
- *new zealand*
- *natural forests*
- *hybrid poplar*
- *commodity standard*
- *fault detection*
- *land management*
- *dual fuel engine*
- *anaerobic*
- *nitrous*
- *flue gas*
- *transitional dynamics*
- *pyrolysis*
- *external resistance*
- *resource*
- *uranium mining*
- *sulphur*
- *expected life*
- *depletion*
- *economic performance*
- *geoscience*
- *regional economy*
- *silicon wafers*
- *sewage*
- *biofilm*
- *electrode*
- *heated*
- *terminal voltage*
- *hydrolysis*
- *economic effects*
- *environmental protection agency*
- *semi-empirical models*
- *rice hulls*
- *extreme weather*
- *international space station*
- *member state*
- *forest resources*
- *financial incentives*
- *depletion rate*
- *policy implementation*
- *agricultural lands*
- *overlapping generations*
- *galapagos archipelago*
- *component testing*
- *creation of credit*
- *land resources*
- *caspian sea*
- *role of particles*
- *international agreements*
- *wine industry*
- *welding*
- *local residents*
- *life-cycle*
- *non-governmental organizations*
- *government of canada*
- *economic policy*
- *project manager*
- *balanced growth*
- *enthalpy*
- *battery charger*
- *architectural elements*
- *national bureau of economic research*
- *human capital*
- *plants and animals*
- *create jobs*
- *miles per hour*
- *water conservation*
- *semiconductor material*
- *vortex*
- *capital accumulation*
- *economic cooperation*
- *central shaft*
- *irradiance*
- *technology transfer*
- *strong winds*
- *third world*
- *prime minister*
- *launch vehicle*
- *full-sized*
- *drilling operations*
- *electronic components*
- *vibration control*
- *subcontractor*
- *emergency management*
- *yucca mountain*
- *chemical engineer*
- *hydraulic*
- *economic costs*

- *world bank*
- *tax subsidies*
- *atmospheric sciences*
- *geosynchronous orbit*
- *electroplating*
- *metal recovery*
- *construction materials*
- *endogenous growth*
- *optimal taxation*
- *resources management*
- *asset management*
- *biocatalyst*
- *federal reserve*
- *joint management*
- *strategic action*
- *taxpayer*
- *grounding*
- *forest service*
- *virgin islands*
- *fire protection association*
- *human relations*
- *forest products*
- *stand-alone*
- *labor input*
- *fall in price*
- *policy research*
- *load factor*
- *united states*
- *fahrenheit*
- *high-pressure*
- *federal agencies*
- *growth path*
- *systems engineering*
- *foreign policy*
- *corrosion*
- *biological attack*
- *donate*
- *permafrost*
- *consumer behaviour*
- *economic dynamics*
- *climatology*
- *electrodes*
- *fort lauderdale*
- *sealed*
- *economic information*
- *gross domestic product*
- *economic growth*
- *storm damage*
- *reactor safety*
- *fiscal year*
- *economic recovery*
- *pacific coast of north*
- *capital expenditure*
- *private investment*
- *high-altitude*
- *decision-making*
- *transaction cost economics*
- *enquiries*
- *rock formations*
- *capital flow*
- *access key*
- *global population*
- *incubator*
- *equilibrium path*
- *long-run growth*
- *water use*
- *secretary of state*
- *heat wave*
- *cross flow*
- *friction loss*
- *inflation rate*
- *cost-effectiveness*
- *millions of years*
- *international law*
- *repeal*
- *master of engineering*
- *data integration*
- *structural materials*
- *spring water*
- *uranium*
- *regulating*
- *information resources*
- *user area*
- *control circuit*
- *great plains*
- *business and industry*
- *ceramic*
- *physical capital*
- *weather stations*
- *environmental biotechnology*
- *cost of equipment*
- *choice experiments*
- *chemical safety*
- *fast-growing*
- *eastern european countries*
- *contact dr*
- *response strategies*
- *light-weight concrete*
- *interim report*
- *subsidiary*
- *international network*
- *public economics*
- *competitive edge*
- *nonlinear control*
- *annual mean*
- *production function*
- *rinse water*
- *earth science*
- *financial analysis*
- *effluents*
- *improved model*
- *public benefits*
- *non-stop*
- *amortization*
- *economic research*
- *freedom of information*
- *franchise tax*
- *concerted*
- *poultry litter*
- *low-income families*
- *glass fibres*
- *reverse osmosis*
- *hydrology*
- *government documents*
- *publication number*
- *educational institutions*
- *guaranteed loan*
- *co-operative*
- *absolute level*
- *home shopping*
- *organic substrate*
- *service provider*
- *education programs*
- *e-commerce*
- *canadian association*
- *interest rate*
- *chemical oxygen demand*
- *growth trends*
- *beryllium disease*
- *optimization*
- *rotary motion*
- *plate tectonics*
- *investor*
- *high-energy*
- *china*
- *actuators*
- *protected by copyright*
- *tk electrical engineering electronics nuclear engineering*
- *matrix composites*
- *intensity effects*
- *index terms*
- *manifold*
- *molten rock*
- *wood fiber*
- *antitrust*
- *equivalent model*

- *authorised*
- *organic film*
- *aquifer*
- *graphite*
- *enrolment*
- *canary*
- *safe environment*
- *ieee trans*
- *groundwork*
- *data center*
- *ceramic matrix composites*
- *organizational development*
- *agricultural research service*
- *gathering information*
- *endogenous*
- *aisi 316l stainless steel grade type*
- *fatigue life*
- *metal cap*
- *inertia*
- *logit model*
- *attitudinal factors*
- *tasmania*
- *information center*
- *fair use*
- *gibraltar strait*
- *attributable*
- *geologic sediments*
- *endogenous variables*
- *option to purchase*
- *sandstone*
- *dinosaurs*
- *open loop control system*
- *bacteria*
- *turbulent*
- *support scheme*
- *fatality rates*
- *subscribed*
- *cash flow analysis*
- *subsurface*
- *physical education*
- *cost functions*
- *strait*
- *de janeiro*
- *r&d*
- *distance learning*
- *rayleigh distribution*
- *chronic beryllium disease*
- *replenish*
- *ress*
- *western china*
- *pole-placement*
- *qualitative interviews*
- *structural equation model*
- *composite repair*
- *biosphere*
- *adhesives*
- *top plate*
- *multi-stage*
- *explicit formula*
- *merger*
- *power ratio*
- *intermittent*
- *structural equation*
- *rotational*
- *accretion*
- *circulation features*
- *ultimate load*
- *alternating*
- *polarization*
- *dynamical transition*
- *proton*
- *milieu*
- *pressure decline*
- *fractures*
- *450*
- *short-term variability*
- *internal haemorrhage*
- *pulmonary haemorrhage*
- *pulmonary lesions*
- *barotrauma*
- *visual factors*
- *conservation equation*
- *risk profile*
- *edgewise*

A.6 Growth rate analysis

Ranked growth rates for 1,343 renewable energy terms, normalized ratio and normalized exponential rate:

Normalized ratio:		Normalized exponential rate:	
space technologies	14.6	*renewable energy policy harmonisation*	4.271
dinosaurs	13.57	*optimal renewable energy mathematical model*	4.161
challenge 2008	7.643	*ecologically optimized transport systems*	4.051
book reviews	5.207	*energy-saving technical progress*	3.167
renewable energy incentives	4.638	*tourist ventures*	3.131
access key	4.606	*centre of appropriate technology*	3.084
nanotechnology	4.066	*load powering*	2.992
microhydro	3.978	*renewable energy promotion strategies*	2.992
pv-compact	3.948	*accelerated depreciation plan*	2.923
donate	3.706	*renewable energy development strategies*	2.813
petroleum resources	3.589	*renewable energy generation cost*	2.784
environmentalism	3.402	*optimal renewable energy mathematical (orem) model*	2.784
european islands	3.189	*emission control policy*	2.754
editorials	3.156	*household energy situation*	2.715
testing phase	3.022	*carbon dioxide discharges*	2.664
chemical safety	2.87	*hybrid diesel systems*	1.944
earth science	2.677	*decentralized power systems penetration*	1.783
vibration control	2.405	*intermittent renewable energy penetration*	1.783
chronic beryllium disease	2.342	*decarbonised world*	1.575
beryllium disease	2.255	*biogas electricity conversion*	1.14
food technology	2.205	*domestic technology manufacturing*	1.14
energy politics	2.182	*renewable energy engineering curriculum*	1.14
articles	2.114	*renewable energy engineering system*	1.14
alternative fuels	1.973	*workshop in a box program*	1.14
earth climate change	1.949	*dinosaurs*	1.14
global warming	1.93	*commercial energy scene*	1.071
wind systems	1.892	*indian national programme*	1.071
energy technology	1.886	*market adoptions*	1.071
biofuel	1.744	*space technologies*	1.033
environmental engineering	1.664	*book reviews*	0.874
greenhouse effect	1.611	*nanotechnology*	0.609
carbon cap-and-trade	1.603	*access key*	0.599
wind energy	1.573	*challenge 2008*	0.592
carbon emissions	1.568	*pv-compact*	0.501
biofuels	1.567	*microhydro*	0.495
failure analysis	1.546	*petroleum resources*	0.478

biofuel energy	1.495	***donate***	0.457
renewable energy research	1.411	***vibration control***	0.439
research output	1.408	***chronic beryllium disease***	0.424
solar water	1.394	***editorials***	0.423
marine renewable energy	1.39	***beryllium disease***	0.402
offshore wind turbine	1.329	***renewable energy incentives***	0.368
ignalina nuclear power plant	1.306	***alternative fuels***	0.363
supercomputing	1.291	***chemical safety***	0.356
volcano	1.289	***earth science***	0.349
estonia	1.237	***articles***	0.316
continuous energy supply	1.215	***global warming***	0.315
gibraltar strait	1.215	***environmentalism***	0.314
legal environment	1.195	***wind systems***	0.299
publications	1.181	***energy technology***	0.297
renewable energies	1.175	***greenhouse effect***	0.292
geyser	1.159	***european islands***	0.29
geoscience	1.157	***wind energy***	0.265
solar energy	1.151	***testing phase***	0.224
solar energy applications	1.14	***supercomputing***	0.22
solar power	1.138	***biofuel***	0.214
contact dr	1.132	***solar water***	0.186
low carbon	1.126	***renewable energy research***	0.181
human genome	1.126	***biofuels***	0.178
renewable energy investment	1.109	***food technology***	0.167
geothermal systems	1.109	***energy politics***	0.163
wave energy	1.098	***earth climate change***	0.158
bioresources	1.095	***offshore wind turbine***	0.152
china	1.073	***volcano***	0.136
reviews	1.072	***estonia***	0.135
sustainability	1.07	***environmental engineering***	0.127
hydrogen fuel	1.062	***carbon emissions***	0.121
fuel cells	1.056	***geoscience***	0.114
australia	1.056	***renewable energy investment***	0.112
media relations	1.043	***reviews***	0.11
fossil fuels	1.043	***marine renewable energy***	0.109
national security	1.028	***fossil fuels***	0.103
solar city	1.014	***biofuel energy***	0.089
energy	1	***fuel cells***	0.085
renewable energy	1	***carbon cap-and-trade***	0.08
alternative energy	0.999	***publications***	0.076
thermonuclear power	0.987	***climate policy***	0.073
industrial technologies	0.984	***solar power***	0.07
india	0.969	***australia***	0.069
power utility	0.963	***emissions trading***	0.067
pakistan	0.962	***china***	0.065
pump systems	0.961	***failure analysis***	0.057
turkey	0.951	***solar house***	0.055
usa	0.95	***contact dr***	0.055

electric vehicles	0.945	*human genome*	0.049
wind farm	0.943	*sustainability*	0.046
national parks	0.941	*low carbon*	0.044
renewable energy standards	0.934	*pakistan*	0.038
wind	0.934	*wind*	0.036
uk	0.923	*usa*	0.032
spain	0.92	*uk*	0.024
climate policy	0.917	*wave energy*	0.022
solar house	0.913	*renewable energies*	0.02
emerging renewable energy markets	0.911	*india*	0.018
hybrid solar-wind system	0.911	*turkey*	0.018
tax credit	0.887	*spain*	0.018
greenhouse gases	0.887	*hydrogen fuel*	0.016
solar electricity	0.886	*solar energy*	0.015
transportation	0.884	*solar city*	0.013
renewable energy materials	0.884	*geyser*	0.012
ethanol	0.879	*alternative energy*	0.009
oil	0.876	*greenhouse gases*	0.002
mexico	0.871	*ignalina nuclear power plant*	0.001
environmental policy	0.852	*1.27e15 j*	0
internet	0.852	*ad 2020-2021*	0
emissions trading	0.845	*asean economic sector*	0
environmental education	0.844	*australian cooperative research centre for renewable energy research program*	0
energy research	0.841	*co2 emission control policy*	0
energy policy	0.835	*coherent renewable energy policy*	0
greece	0.827	*cost/efficiency ratio minimisation*	0
wind power plant	0.825	*crude oil supplies reduction*	0
welding	0.822	*domestic renewable energy generators*	0
alternative fuel	0.813	*electric power policy frameworks*	0
remote power generation	0.812	*electrical power facility (epf)*	0
letters	0.806	*energy*	0
community	0.801	*energy density utilization*	0
bio energies	0.797	*energy-price signals reformation*	0
wind turbines	0.795	*environmental friendly sustainable development*	0
wind power	0.784	*finance pre-feasibility studies*	0
malaysia	0.784	*financing risk premiums reduction*	0
biodiesel	0.782	*flexible photo-voltaic solar cell*	0
taiwan	0.779	*fossil fuel politics*	0
actuators	0.776	*fuel for thought strategies*	0
quality control	0.774	*geothermal resources -- environmental aspects -- oregon -- prineville region*	0
hydrogen	0.769	*global renewable energy movement*	0
mechanical engineering	0.765	*high activity subjects*	0
lead-acid battery	0.763	*intrinsic technology problem*	0
renewable energy development bill	0.759	*kirklees councils*	0
greenhouse gas emissions	0.754	*low activity subjects*	0
washington	0.753	*marginalized mountain population*	0

government	0.753	*multilevel search engines*	0
energy efficiency	0.747	*natural gas supplies reduction*	0
maldives	0.743	*nonpooled electricity trading*	0
galapagos archipelago	0.74	*petroleum crude oil consumption*	0
light energy	0.731	*power plant financing process*	0
environmentalists	0.73	*project datums*	0
greenhouse gas	0.727	*public lands -- oregon -- prineville region -- management*	0
blades	0.727	*recursive linear planning model*	0
plate tectonics	0.721	*recursive linear program*	0
renewable energy potential	0.721	*regional power market competition*	0
solar wind	0.715	*regulated rural energy concessions*	0
wind turbine	0.714	*renewable energy*	0
fuel standards	0.713	*renewable energy generating equipments*	0
germany	0.71	*renewable energy product purchase*	0
building energy	0.709	*renewable energy systems reliability*	0
mechatronics	0.706	*renewable energy trading experience*	0
argentina	0.704	*renewable facilites*	0
wind farms	0.702	*ring-fenced markets*	0
freedom of information	0.699	*social acceptance variation*	0
energy industry	0.697	*socio economic optimal renewable energy model*	0
solar power plant	0.692	*sustained renewable energy deployment*	0
renewable energy markets	0.691	*tax revenue method*	0
renewable energy sector	0.682	*tertiary renewable energy education*	0
belgium	0.682	*tertiary trained engineers*	0
carbon dioxide emissions	0.679	*tertiary trained policy makers*	0
climatology	0.677	*tertiary trained scientists*	0
planning	0.677	*time 2020 year to 2021 year*	0
solar concentrator	0.675	*uniform green power market mode*	0
nuclear science	0.673	*windfarm market*	0
electrical engineering	0.672	*media relations*	0
energy model	0.672	*wind farm*	-0.001
nuclear data	0.665	*electric vehicles*	-0.002
greenhouse gas reduction	0.661	*geothermal systems*	-0.004
coal-fired power plants	0.659	*wind power plant*	-0.005
new zealand	0.657	*environmental policy*	-0.007
insulation	0.656	*national security*	-0.013
efficiency	0.655	*mexico*	-0.021
course structure	0.654	*emerging renewable energy markets*	-0.022
maps	0.653	*oil*	-0.023
energy studies	0.645	*bioresources*	-0.024
latvia	0.644	*energy research*	-0.032
undergraduate education	0.642	*industrial technologies*	-0.034
energy storage	0.639	*carbon dioxide emissions*	-0.037
output-based allocation	0.638	*malaysia*	-0.038
ipcc	0.636	*national parks*	-0.04

energy generation	0.635	***transportation***	-0.04
ocean energy	0.634	***wind turbines***	-0.044
distance learning	0.631	***greece***	-0.044
fuel economy	0.629	***research output***	-0.045
solar cell	0.627	***solar water heating systems***	-0.045
methane	0.626	***hybrid power systems***	-0.051
computational science	0.625	***legal environment***	-0.052
thermal systems	0.625	***alternative fuel***	-0.057
projects	0.622	***solar energy applications***	-0.057
vice president	0.617	***continuous energy supply***	-0.058
energy farming	0.615	***taiwan***	-0.062
clean coal	0.614	***ethanol***	-0.062
alternative energy source	0.611	***wind power***	-0.063
renewable energy supply (res)	0.607	***pump systems***	-0.063
state of michigan	0.607	***galapagos archipelago***	-0.063
substitute cost	0.607	***solar electricity***	-0.065
turbines	0.605	***environmental education***	-0.065
water heaters	0.605	***power utility***	-0.07
physical education	0.604	***belgium***	-0.07
japan	0.604	***welding***	-0.071
geothermal energy	0.603	***germany***	-0.075
grounding	0.603	***solar power plant***	-0.077
energy systems engineering	0.6	***environmentalists***	-0.077
terrorism	0.598	***renewable energy materials***	-0.078
harnessing	0.597	***community***	-0.079
wave power	0.595	***washington***	-0.081
hydrogen vehicles	0.595	***biodiesel***	-0.081
solar hydrogen	0.593	***argentina***	-0.082
scientific challenges	0.593	***quality control***	-0.083
megawatts	0.592	***solar wind***	-0.083
world leaders	0.59	***physical education***	-0.084
product design	0.589	***government***	-0.085
intergovernmental panel on climate change	0.588	***hydrogen***	-0.088
environmental problems	0.586	***lead-acid battery***	-0.088
electricity	0.585	***gibraltar strait***	-0.091
taxpayer	0.585	***mechanical engineering***	-0.091
renewable energy planning	0.58	***tax credit***	-0.092
energy security	0.578	***energy policy***	-0.095
renewable energy utilisation	0.577	***renewable energy standards***	-0.097
volcanic zone	0.577	***wind turbine***	-0.099
northern queensland	0.576	***energy efficiency***	-0.1
scale-up	0.575	***bio energies***	-0.101
investor	0.574	***new zealand***	-0.103
intergovernmental panel	0.574	***building energy***	-0.106
commercialization	0.574	***light energy***	-0.106
oil prices	0.573	***methane***	-0.107
clean fuels	0.573	***blades***	-0.108

tidal power	0.573	***letters***	-0.108
home energy	0.571	***wind farms***	-0.111
r&d	0.571	***renewable energy sector***	-0.115
r&d support	0.571	***course structure***	-0.118
renewable energy r&d	0.571	***ocean energy***	-0.119
gasoline	0.568	***fuel standards***	-0.119
renewable energy industry	0.567	***energy industry***	-0.119
heating	0.565	***remote power generation***	-0.121
geothermal	0.565	***renewable energy utilisation***	-0.123
roaring forties	0.564	***renewable energy markets***	-0.123
atmospheric carbon dioxide	0.559	***greenhouse gas emissions***	-0.124
global energy demand	0.559	***maldives***	-0.125
renewable energy projects	0.555	***greenhouse gas***	-0.129
energy utilities	0.555	***renewable energy development bill***	-0.13
energy efficient	0.555	***nuclear data***	-0.131
legislation	0.554	***mechatronics***	-0.133
renewables	0.554	***planning***	-0.135
sustainable future	0.552	***plate tectonics***	-0.135
rotor blades	0.55	***electrical power test***	-0.139
senior vice president	0.547	***small wind turbines***	-0.143
semi-empirical models	0.547	***japan***	-0.146
energy certificate	0.546	***accelerator***	-0.149
hydrology	0.545	***efficiency***	-0.15
sustainable development	0.544	***geothermal energy***	-0.15
ocean wave energy	0.542	***latvia***	-0.151
investment	0.539	***western australia***	-0.152
engineering program	0.538	***turbines***	-0.153
sustainable energy	0.538	***social studies***	-0.156
buildings	0.536	***wind excitation***	-0.157
solar thermal	0.534	***energy storage***	-0.157
domestic buildings	0.534	***human genome project***	-0.158
alarm limit	0.531	***freedom of information***	-0.158
australian renewable energy sector	0.531	***storage technology***	-0.158
photovoltaic generation system	0.531	***tasmania***	-0.159
reliability constraints	0.531	***electrical engineering***	-0.16
carbon sequestration	0.528	***data center***	-0.16
high-altitude	0.528	***projects***	-0.16
lighting	0.527	***power system design***	-0.16
growing economy	0.525	***distance learning***	-0.161
policies	0.523	***clean coal***	-0.161
northern australia	0.521	***fuel economy***	-0.161
central and eastern europe	0.52	***climatology***	-0.164
batch reactors	0.519	***product design***	-0.164
biological attack	0.517	***maps***	-0.165
propulsion	0.517	***thermal systems***	-0.167
notes	0.517	***sustainable development***	-0.168
renewable energy plan	0.513	***energy generation***	-0.168
oil reserves	0.513	***rotor blades***	-0.168

rock formations	0.512	***electric vehicle***	-0.17
solar hot water system	0.511	***bangladesh***	-0.17
best practices	0.511	***actuators***	-0.17
global security	0.511	***terrorism***	-0.171
internationally focused	0.511	***tidal power***	-0.172
co2	0.511	***energy studies***	-0.172
fuel cell	0.511	***renewable energy planning***	-0.173
western australia	0.51	***energy systems engineering***	-0.176
electric vehicle	0.507	***investor***	-0.176
incubator	0.507	***coal-fired power plants***	-0.176
hot water systems	0.506	***computational science***	-0.177
turbine	0.506	***renewable energy projects***	-0.177
virgin islands	0.505	***water heaters***	-0.179
renewable energy policies	0.505	***internet***	-0.179
solar pv	0.504	***megawatts***	-0.179
transportation fuels	0.501	***renewable energy statistics***	-0.179
fossil fuel economy	0.501	***vice president***	-0.182
renewable energy statistics	0.501	***nuclear science***	-0.182
fuel efficiency	0.5	***solar water heating***	-0.182
photovoltaic energy	0.5	***renewable energy industry***	-0.182
small wind turbines	0.5	***solar cell***	-0.183
energy extraction	0.5	***sustainable future***	-0.183
accelerator	0.499	***wave power***	-0.185
sustainable buildings	0.498	***renewables***	-0.186
power systems engineering	0.495	***geothermal***	-0.186
bioenergy trade	0.494	***public affairs***	-0.187
pelton wheel	0.494	***solar pv***	-0.187
financing	0.493	***insulation***	-0.188
public affairs	0.493	***electricity***	-0.19
energy unit	0.49	***renewable energy plan***	-0.191
tidal wave	0.489	***solar concentrator***	-0.193
general electric	0.489	***efficient lighting***	-0.193
atmospheric sciences	0.487	***energy efficient***	-0.196
bangladesh	0.486	***heating***	-0.197
auxiliary service	0.486	***solar thermal***	-0.197
hindu kush himalayan region	0.486	***buildings***	-0.2
tidal channel	0.486	***renewable energy potential***	-0.201
data center	0.484	***lighting***	-0.202
cooling	0.483	***home energy***	-0.202
human genome project	0.483	***environmental problems***	-0.203
efficient lighting	0.483	***greenhouse gas reduction***	-0.203
storage technology	0.481	***state of michigan***	-0.203
tasmania	0.481	***alarm limit***	-0.203
built environment	0.481	***legislation***	-0.205
global climate change	0.48	***nuclear safety***	-0.205
murdoch university	0.477	***renewable energy policies***	-0.205
social studies	0.476	***grounding***	-0.205
renewable energy generation	0.475	***ipcc***	-0.207

nepal	0.474	***lead-acid batteries***	-0.208
bioenergy	0.472	***energy security***	-0.209
power generation	0.472	***harnessing***	-0.21
energy potential	0.472	***gasoline***	-0.211
eu energy performance of buildings directive	0.471	***power systems***	-0.211
consumer products	0.47	***northern queensland***	-0.211
energy applications	0.469	***clean fuels***	-0.212
solar heating	0.468	***solar hydrogen***	-0.213
energy management	0.466	***hybrid solar-wind system***	-0.213
balanced energy-mix	0.466	***world leaders***	-0.214
european union	0.465	***livermore national laboratory***	-0.214
pv	0.464	***solar heating***	-0.214
renewable energy financing	0.464	***turbine***	-0.215
energy crisis	0.464	***energy farming***	-0.216
hydropower	0.464	***hot water systems***	-0.216
uranium	0.463	***built environment***	-0.217
building regulations	0.463	***oil prices***	-0.217
thermal power	0.463	***water pumping***	-0.218
biomass	0.462	***fuel cell***	-0.219
nuclear research	0.462	***roaring forties***	-0.219
gas supply	0.461	***renewable energy supply (res)***	-0.219
reducing greenhouse gas emissions	0.457	***investment***	-0.22
executive director	0.456	***bioenergy***	-0.22
bachelor degree program	0.456	***intergovernmental panel on climate change***	-0.221
china-india cooperation	0.456	***senior vice president***	-0.221
renewable energy centres	0.456	***sustainable energy***	-0.224
renewislands	0.456	***international law***	-0.228
satellite-based sensors	0.456	***intergovernmental panel***	-0.23
hybrid development	0.456	***alternative energy source***	-0.231
western flank	0.456	***carbon sequestration***	-0.231
aerodynamics	0.454	***murdoch university***	-0.231
hydrogen energy	0.451	***energy model***	-0.232
energy economics	0.451	***solar hot water system***	-0.233
nuclear safety	0.451	***thermonuclear power***	-0.233
power systems	0.449	***electric power systems***	-0.233
dynamic testing	0.449	***ocean wave energy***	-0.235
energy technologies	0.447	***taxpayer***	-0.236
public participation	0.446	***global security***	-0.236
cooking	0.445	***efficient appliances***	-0.236
innovative business models	0.445	***hybrid development***	-0.237
environmental contamination	0.444	***undergraduate education***	-0.237
energy sources	0.443	***atmospheric carbon dioxide***	-0.238
sequencing batch reactors	0.442	***energy utilities***	-0.238
united nations environment programme	0.442	***central and eastern europe***	-0.238
solar water heating systems	0.439	***cooking***	-0.24
global markets	0.438	***commercialization***	-0.241
government mandates	0.438	***photovoltaic***	-0.242

job losses	0.437	***policies***	-0.244
renewable energy technology research	0.437	***hydrology***	-0.245
tidal current	0.435	***scale-up***	-0.245
case studies	0.434	***propulsion***	-0.245
ocean breezes	0.434	***pv***	-0.245
pumping	0.434	***northern australia***	-0.246
fahrenheit	0.433	***high-altitude***	-0.247
photovoltaic	0.431	***photovoltaic energy***	-0.247
power supply	0.431	***energy certificate***	-0.248
investments financing	0.43	***energy management***	-0.249
bio energy	0.43	***co2***	-0.25
great plains	0.429	***power systems engineering***	-0.25
research and developments	0.429	***cooling***	-0.251
renewable energy systems	0.428	***investments financing***	-0.251
energy systems	0.428	***project development***	-0.251
technology transfer	0.427	***american council***	-0.252
wind system	0.426	***photovoltaic systems***	-0.254
maintenance	0.426	***nuclear technology***	-0.254
energy policies	0.425	***global energy demand***	-0.254
northern south australia	0.425	***energy-efficient lighting***	-0.255
renewable energy curriculum	0.425	***biological attack***	-0.257
wind power generation	0.424	***hydrogen vehicles***	-0.258
renewable portfolio standards	0.423	***volcanic zone***	-0.259
barrels per day	0.423	***information service***	-0.259
industrial ecology	0.423	***fuel efficiency***	-0.259
renewable energy technology	0.422	***biomass***	-0.261
plasma physics	0.422	***pumping***	-0.261
aquifers	0.422	***solar systems***	-0.262
future energy	0.421	***notes***	-0.263
fuel storage	0.42	***wind system***	-0.263
international law	0.419	***engineering program***	-0.263
production tax	0.418	***r&d support***	-0.265
islands	0.418	***renewable energy r&d***	-0.265
renewable energy sources	0.418	***r&d***	-0.265
electricity market reforms	0.418	***water heating***	-0.265
global renewable energy industry	0.418	***nepal***	-0.266
linear generator	0.418	***renewable energy centres***	-0.267
environmental groups	0.417	***hydropower***	-0.268
world bank	0.416	***aerodynamics***	-0.269
power stations	0.415	***oil reserves***	-0.27
energy usage	0.415	***energy systems***	-0.27
joint ventures	0.415	***european union***	-0.271
diesel	0.413	***scientific challenges***	-0.271
record search	0.412	***power generation***	-0.272
research and development	0.411	***ownership system***	-0.272
global issues	0.411	***materials technology***	-0.273
renewable energy technologies	0.411	***growing economy***	-0.273
environmental issues	0.411	***uranium***	-0.274

solar energy conversion	0.41	***global climate change***	-0.275
solar energy utilization	0.409	***global renewable energy industry***	-0.275
power electronics	0.409	***world bank***	-0.276
onshore	0.409	***tidal wave***	-0.276
aquifer	0.406	***energy supply security considerations***	-0.277
mitigation	0.406	***energy potential***	-0.278
efficient appliances	0.405	***renewable energy systems***	-0.278
wind facilities	0.404	***consumer products***	-0.279
energy prices	0.404	***sustainable buildings***	-0.279
environment programme	0.403	***financing***	-0.282
tariff	0.402	***internationally focused***	-0.282
emissions control	0.402	***virgin islands***	-0.282
triple point	0.401	***general electric***	-0.282
hybrid power systems	0.401	***islands***	-0.283
nuclear technology	0.401	***batch reactors***	-0.284
nuclear weapons	0.4	***fossil fuel economy***	-0.286
west texas	0.399	***energy crisis***	-0.287
renewable energy policy	0.399	***thermal power***	-0.287
microforms	0.399	***substitute cost***	-0.288
technology push	0.399	***energy economics***	-0.289
renewable sources	0.398	***power electronics***	-0.29
electrical and computer engineering (ece)	0.398	***incubator***	-0.29
power system design	0.397	***technology transfer***	-0.291
senegal	0.397	***solar collectors***	-0.292
state government	0.397	***energy sources***	-0.292
enhanced greenhouse effect	0.396	***heating system***	-0.293
environmentally friendly alternatives	0.395	***domestic buildings***	-0.293
energy savings	0.395	***maintenance***	-0.293
strong winds	0.393	***energy extraction***	-0.294
solar water heating	0.393	***transportation fuels***	-0.294
create jobs	0.392	***best practices***	-0.294
energy market	0.391	***atmospheric sciences***	-0.295
biogas	0.391	***renewable energy technology***	-0.295
cell technology	0.391	***environmental groups***	-0.295
energy resource	0.39	***energy policies***	-0.297
sustainable energy sources	0.39	***output-based allocation***	-0.298
heat energy	0.389	***executive director***	-0.298
low-income families	0.389	***energy technologies***	-0.299
organizational development	0.388	***sustainable growth***	-0.299
solar collectors	0.388	***reliability constraints***	-0.299
geothermal activity	0.388	***wind power generation***	-0.299
storage technologies	0.387	***nuclear research***	-0.301
energy system analysis	0.387	***photovoltaics***	-0.301
conservation of energy	0.387	***bioenergy trade***	-0.301
information service	0.387	***innovative business models***	-0.302
european renewable energy council (erec)	0.386	***fahrenheit***	-0.302
project development	0.385	***senegal***	-0.303
storm damage	0.384	***european renewable energy council***	-0.303

		(erec)	
framework programme	0.384	*nuclear weapons*	-0.304
optimization	0.384	*onshore*	-0.304
renewable energy investments	0.384	*renewable energy generation*	-0.304
tidal stream	0.383	*reducing greenhouse gas emissions*	-0.306
commercialisation	0.383	*tidal channel*	-0.307
heating system	0.383	*renewable energy sources*	-0.307
renewable energy education	0.383	*renewable energy education*	-0.307
bio-energy systems	0.383	*gas supply*	-0.307
tidal energy	0.383	*government mandates*	-0.308
electricity grid	0.382	*public participation*	-0.308
government support	0.382	*tidal current*	-0.308
energy sector	0.382	*balanced energy-mix*	-0.308
solar systems	0.382	*research and development*	-0.309
wind loading	0.382	*production tax*	-0.309
heat pump	0.381	*rock formations*	-0.31
energy demand	0.381	*case studies*	-0.31
alloy surfaces	0.38	*systems research*	-0.311
financing processes	0.38	*fuel cell systems*	-0.311
renewable energy assessment	0.38	*solar thermal energy*	-0.311
sustainable growth	0.379	*pelton wheel*	-0.311
electric grid	0.379	*energy prices*	-0.312
water pumping	0.379	*renewable energy financing*	-0.312
energy conservation	0.379	*energy unit*	-0.313
reliability	0.378	*research and developments*	-0.313
environmental impact assessment	0.378	*job losses*	-0.314
electric range	0.378	*tidal energy*	-0.314
livermore national laboratory	0.377	*building regulations*	-0.314
non-renewable energy	0.377	*conservation of energy*	-0.314
central asia	0.377	*future energy*	-0.315
design for sustainability	0.376	*ministry of nonconventional energy sources*	-0.315
american council	0.376	*energy products*	-0.316
northern territory	0.375	*hydrogen energy*	-0.316
energy economics and policy	0.375	*plasma physics*	-0.316
water heating	0.375	*world wide web*	-0.316
social issues	0.371	*social issues*	-0.317
office of basic energy sciences	0.371	*eu energy performance of buildings directive*	-0.317
total energies	0.371	*sequencing batch reactors*	-0.318
impact on the environment	0.371	*strong winds*	-0.318
installed capacity	0.371	*aquifers*	-0.318
fossil fuel	0.37	*renewable energy technology research*	-0.319
power sector	0.37	*power supply*	-0.32
materials technology	0.369	*diesel*	-0.32
state government agencies	0.368	*renewable energy investments*	-0.321
great basin	0.365	*tidal stream*	-0.321
renewable energy development	0.365	*power stations*	-0.322
power quality	0.365	*resource assessment*	-0.322

innovation systems	0.363	***after sales service***	-0.322
employment opportunities	0.362	***emissions control***	-0.324
fission	0.361	***solar thermal systems***	-0.324
energy policy act	0.361	***renewable energy policy***	-0.324
standardization	0.361	***water turbine***	-0.325
technological development	0.36	***energy applications***	-0.326
materials handling	0.359	***technology push***	-0.327
energy mix	0.358	***environmental issues***	-0.328
ecodesign	0.357	***triple point***	-0.328
energy-efficient lighting	0.357	***renewable energy curriculum***	-0.328
fort collins	0.356	***great plains***	-0.329
electroplating	0.355	***bio-energy systems***	-0.33
office of science	0.355	***renewable portfolio standards***	-0.33
solar heating system	0.355	***solar energy conversion***	-0.331
turbulent	0.354	***united nations environment programme***	-0.331
renewable energy commercialization	0.354	***ecodesign***	-0.333
environmental protection agency	0.354	***renewislands***	-0.334
energy consumption	0.353	***barrels per day***	-0.335
coal liquefaction	0.353	***industrial ecology***	-0.335
secretary of state	0.352	***cell technology***	-0.335
energy products	0.352	***heating systems***	-0.336
world wide web	0.351	***heat pump***	-0.336
tax revenue	0.351	***northern south australia***	-0.337
turbine blades	0.351	***standardization***	-0.337
hydraulic	0.35	***renewable energy assessment***	-0.339
school environment	0.35	***auxiliary service***	-0.34
transportation sector	0.35	***environmental systems***	-0.34
photovoltaics	0.35	***global issues***	-0.34
fiscal year	0.349	***geothermal power***	-0.34
hot springs	0.349	***bio energy***	-0.34
fuel power	0.348	***renewable energy technologies***	-0.34
government incentives	0.348	***land management***	-0.341
energy use	0.348	***environmental contamination***	-0.342
renewable energy (re)	0.348	***global markets***	-0.343
participation	0.348	***photovoltaic system***	-0.343
environmental technology	0.347	***solar water heaters***	-0.344
coal research	0.346	***renewable sources***	-0.344
hybrid renewable energy systems	0.346	***energy savings***	-0.344
land management	0.346	***energy sector***	-0.344
renewable energy law	0.346	***western flank***	-0.345
energy audit	0.345	***industrial energy***	-0.346
ownership system	0.345	***cell energy***	-0.347
energy resources	0.345	***energy market***	-0.348
renewable energy resources	0.344	***mitigation***	-0.349
environmental systems	0.344	***energy usage***	-0.349
sustainable energy source	0.344	***storm damage***	-0.349
stand-alone	0.343	***semi-empirical models***	-0.35
construction materials	0.343	***sustainable energy sources***	-0.352

energy capacity	0.343	*microforms*	-0.353
fuel cell systems	0.342	*environmentally friendly alternatives*	-0.353
lead-acid batteries	0.342	*renewable energy development*	-0.353
scheikunde	0.342	*fort collins*	-0.354
water research	0.341	*commercialisation*	-0.354
drilled	0.34	*international network*	-0.355
sources of energy	0.34	*optimization*	-0.355
developing countries	0.339	*environment programme*	-0.356
polluting	0.339	*energy conservation*	-0.356
refurbishment projects	0.339	*joint ventures*	-0.357
energy development	0.339	*environmental impact assessment*	-0.358
tax incentives	0.338	*energy-generating resources*	-0.358
solar photovoltaic	0.337	*photovoltaic cells*	-0.358
solar thermal systems	0.337	*secretary of state*	-0.358
postgraduate program	0.337	*central asia*	-0.359
global population	0.336	*low-income families*	-0.36
borehole	0.335	*solar heating system*	-0.361
cell energy	0.335	*ocean breezes*	-0.362
thermochemical conversion	0.335	*broad-winged hawk*	-0.362
soil gas	0.335	*create jobs*	-0.362
geothermal power	0.335	*energy resource*	-0.362
energy-generating resources	0.334	*hybrid renewable energy systems*	-0.362
resource assessment	0.334	*design for sustainability*	-0.362
policy makers	0.333	*construction materials*	-0.366
photovoltaic system	0.333	*electricity market reforms*	-0.366
quality of life	0.332	*state government*	-0.367
energy demand increase	0.331	*environmental protection agency*	-0.367
fractures	0.331	*west texas*	-0.367
regional development	0.331	*renewable energy law*	-0.367
geothermal system	0.331	*fuel storage*	-0.367
geothermal sources	0.33	*dynamic testing*	-0.367
government documents	0.33	*alloy surfaces*	-0.368
local residents	0.329	*energy demand*	-0.368
energy saves	0.329	*organisation for economic co-operation and development*	-0.369
building industry	0.328	*electroplating*	-0.369
after sales service	0.328	*environmental pollution*	-0.37
energy exchange	0.327	*solar photovoltaic*	-0.371
renewable energy market	0.327	*renewable energy (re)*	-0.371
development mechanism	0.327	*office of science*	-0.373
petroleum gas	0.327	*electricity grid*	-0.373
coal production	0.326	*non-renewable energy*	-0.373
tidal barrage	0.325	*reliability*	-0.373
economic information	0.325	*linear generator*	-0.374
renewable energy penetration	0.324	*energy mix*	-0.375
educational materials	0.324	*electric range*	-0.375
hot spring	0.324	*electric grid*	-0.375
emissions reduction	0.324	*fossil fuel*	-0.376

scaling up	0.324	*northern territory*	-0.376
industrial energy	0.324	*installed capacity*	-0.376
energy production	0.322	*energy development*	-0.377
energy diversification	0.321	*office of basic energy sciences*	-0.377
little research	0.321	*energy economics and policy*	-0.378
future generations	0.321	*hot springs*	-0.378
residential building	0.32	*international energy agency*	-0.378
photovoltaic systems	0.32	*power sector*	-0.379
heated	0.319	*fission*	-0.38
energy converters	0.319	*electrical and computer engineering (ece)*	-0.38
genoa	0.319	*turbine blades*	-0.381
turbine rotor	0.318	*energy audit*	-0.381
green electricity	0.318	*wind facilities*	-0.382
low-interest loans	0.318	*fractures*	-0.382
hydroelectric power	0.317	*electrodialysis*	-0.382
wind power development	0.317	*aquifer*	-0.382
government subsidies	0.316	*energy use*	-0.383
integrated gasification	0.316	*de janeiro*	-0.383
climate control	0.315	*heat energy*	-0.383
waste heat	0.314	*power quality*	-0.384
aerodynamic	0.314	*storage technologies*	-0.384
solar thermal energy	0.313	*organizational development*	-0.384
intelligent software	0.313	*building industry*	-0.384
strait	0.312	*bachelor degree program*	-0.384
hydroelectric	0.311	*coal research*	-0.385
systems research	0.31	*record search*	-0.385
subsurface	0.31	*impact on the environment*	-0.386
renewable portfolio	0.31	*hindu kush himalayan region*	-0.386
biomass conversion	0.31	*renewable energy market*	-0.386
modelling study	0.31	*employment opportunities*	-0.386
consumer behaviour	0.309	*energy policy act*	-0.387
decision-support system	0.309	*turbulent*	-0.387
calculation tools	0.309	*developing countries*	-0.388
non-renewable	0.307	*distributed energy*	-0.388
geothermal power plants	0.307	*solar lanterns*	-0.388
electric power systems	0.307	*government support*	-0.388
local government	0.306	*renewable energy resources*	-0.388
kilowatts	0.305	*tariff*	-0.389
energia elétrica - qualidade	0.305	*energy resources*	-0.389
fuel poverty	0.305	*energy saves*	-0.39
distributed energy	0.305	*turbine rotor*	-0.39
photovoltaic technology	0.305	*solar energy systems*	-0.391
sustainable use	0.305	*geothermal activity*	-0.391
integrated assessment	0.305	*fiscal year*	-0.392
[jel] q48	0.304	*local residents*	-0.393
desertification problem	0.304	*energy capacity*	-0.393
international renewable energy development	0.304	*enhanced greenhouse effect*	-0.393

tidal elevation	0.304	***hydraulic***	-0.395
production of hydrogen	0.303	***polluting***	-0.396
investment decisions	0.303	***government documents***	-0.396
unemployment rates	0.303	***investment deduction***	-0.398
environmental performance	0.302	***quality of life***	-0.398
high-pressure	0.301	***environmental technology***	-0.399
correlation	0.301	***energy consumption***	-0.399
energy demands	0.301	***fuel poverty***	-0.4
solar photovoltaics	0.3	***fuel power***	-0.4
photovoltaic cells	0.3	***sources of energy***	-0.402
solar water heaters	0.3	***materials handling***	-0.402
business reasons	0.3	***solar energy utilization***	-0.403
external energy	0.299	***participation***	-0.404
building stock	0.299	***green electricity***	-0.404
baltic states	0.299	***intelligent software***	-0.404
solar cooker	0.298	***tax incentives***	-0.405
transport sector	0.298	***plutonium***	-0.405
electricity generation	0.297	***great basin***	-0.406
plutonium	0.297	***framework programme***	-0.406
renewable energy capacity	0.296	***sustainable energy source***	-0.409
wave loading	0.296	***coal liquefaction***	-0.409
geothermal heat pumps	0.296	***high-pressure***	-0.41
international energy agency	0.295	***energy converters***	-0.41
urban settings	0.295	***climate control***	-0.411
policy options	0.295	***school environment***	-0.411
power sources	0.295	***policy makers***	-0.413
business plans	0.294	***stand-alone***	-0.413
lca	0.294	***technological development***	-0.413
education programs	0.293	***drilled***	-0.413
clean development mechanism (cdm)	0.292	***strait***	-0.414
wind generation	0.292	***genoa***	-0.414
domestic production	0.291	***integrated assessment***	-0.414
450	0.291	***heat pumps***	-0.414
regulating	0.291	***solar photovoltaics***	-0.414
renewable energy costs	0.291	***kilowatts***	-0.415
top-down	0.291	***energia elétrica - qualidade***	-0.415
price drops	0.29	***thermochemical conversion***	-0.415
miles per hour	0.29	***wind power development***	-0.416
power market	0.29	***innovation systems***	-0.416
uranium mining	0.289	***regional development***	-0.416
life-cycle	0.289	***solar cooker***	-0.416
energy crops	0.289	***hydroelectric***	-0.417
energy supply	0.289	***borehole***	-0.418
drilling operations	0.288	***total energies***	-0.419
environmental pollution	0.287	***energy production***	-0.419
wind resource assessment	0.287	***transportation sector***	-0.419
central australia	0.287	***power market***	-0.42
tidal power station	0.287	***heated***	-0.42

water turbine	0.287	***water research***	-0.42
technology division	0.287	***renewable portfolio***	-0.421
hydrothermal	0.286	***geothermal heat pumps***	-0.421
u.s. energy demand	0.285	***global population***	-0.422
oil supply	0.285	***tidal barrage***	-0.422
impact energy	0.285	***aerodynamic***	-0.423
sandstone	0.284	***local government***	-0.423
gas plants	0.284	***hydro power***	-0.424
natural ecosystems	0.284	***energy system analysis***	-0.425
federal agencies	0.283	***renewable energy commercialization***	-0.427
energy supplies	0.283	***geothermal power plants***	-0.427
international conferences	0.283	***government incentives***	-0.427
option to purchase	0.283	***energy exchange***	-0.427
offshore environment	0.283	***government subsidies***	-0.428
reservoir simulation	0.282	***wind resource assessment***	-0.428
efficiency measures	0.282	***geothermal system***	-0.428
solar energy systems	0.281	***state government agencies***	-0.429
electrodialysis	0.281	***miles per hour***	-0.43
electricity supply	0.281	***energy demands***	-0.431
national economy	0.28	***wind loading***	-0.431
life cycle assessment	0.28	***economic co-operation***	-0.432
clean coal technologies	0.28	***economic information***	-0.432
general public	0.279	***development mechanism***	-0.432
renewable electricity generation	0.279	***clean coal technologies***	-0.433
wave energy conversion	0.279	***hot spring***	-0.433
decision-making	0.279	***energy diversification***	-0.433
international network	0.279	***calculation tools***	-0.434
heat pumps	0.279	***emissions reduction***	-0.434
solar insolation	0.279	***integrated gasification***	-0.435
deregulation	0.279	***waste heat***	-0.435
technological research	0.278	***educational materials***	-0.436
environmental impacts	0.278	***oil supply***	-0.437
electricity use	0.277	***tax revenue***	-0.437
emission scenarios	0.276	***geothermal sources***	-0.437
heating systems	0.276	***australian renewable energy sector***	-0.438
pump system	0.276	***postgraduate program***	-0.439
coal combustion	0.275	***biomass energy***	-0.44
federal commitment	0.275	***soil gas***	-0.44
hydro power	0.274	***technological research***	-0.44
electricity demand	0.274	***technology division***	-0.442
conventional power	0.273	***future generations***	-0.443
delphi study	0.273	***photovoltaic technology***	-0.443
small-scale renewable energy technologies	0.273	***coal production***	-0.444
transient modelling	0.273	***satellite-based sensors***	-0.444
franchise tax	0.273	***biomass conversion***	-0.445
wind resource	0.273	***low-interest loans***	-0.446
comparative analysis	0.272	***scaling up***	-0.446

fossil fuel electricity generation	0.272	***renewable energy promotion law***	-0.447
human health	0.272	***hydroelectric power***	-0.448
energy planning	0.272	***energy crops***	-0.449
employment growth	0.272	***heat exchangers***	-0.449
hydrogen systems	0.272	***environmental performance***	-0.449
use of hydrogen	0.272	***non-renewable***	-0.449
service provider	0.271	***offshore environment***	-0.45
fall in price	0.271	***refurbishment projects***	-0.451
negative effects	0.271	***little research***	-0.451
deep drilling	0.271	***energy planning***	-0.452
residential buildings	0.271	***energy supplies***	-0.452
tertiary education	0.271	***energy supply***	-0.452
desalinated water	0.27	***subsurface***	-0.452
distributed energy generation	0.269	***energy demand increase***	-0.453
small communities	0.269	***residential building***	-0.453
questionnaire surveys	0.269	***price drops***	-0.453
de janeiro	0.269	***deep drilling***	-0.454
geothermal drilling	0.269	***urban settings***	-0.454
outlays	0.269	***power sources***	-0.455
energy intensive economy	0.269	***transport sector***	-0.455
renewable energy usage	0.268	***electricity generation***	-0.455
ecologically sustainable development	0.268	***international renewable energy development***	-0.456
policy incentives	0.268	***collaborative initiatives***	-0.457
integration system	0.268	***tidal power station***	-0.457
cross-flow	0.268	***baltic states***	-0.457
power development	0.268	***regulating***	-0.457
power resource	0.267	***petroleum gas***	-0.457
government of india	0.266	***investment decisions***	-0.458
investment deduction	0.266	***renewable energy costs***	-0.458
optimal placement	0.264	***unemployment rates***	-0.458
domestic energy	0.264	***education programs***	-0.459
ress	0.263	***dual fuel engine***	-0.459
building energy efficiency	0.263	***hydrothermal***	-0.459
milieu	0.263	***heating water***	-0.459
primary energy source	0.263	***production of hydrogen***	-0.46
dual fuel engine	0.262	***consumer behaviour***	-0.46
power system planning	0.262	***milieu***	-0.46
power production	0.262	***environmental conservation***	-0.46
production systems	0.262	***renewable energy capacity***	-0.461
international cooperation	0.262	***domestic production***	-0.461
aboriginal communities	0.26	***federal agencies***	-0.461
renewable energy promotion law	0.26	***sustainable use***	-0.461
issue date	0.26	***wind generation***	-0.461
spatial planning	0.26	***diffusion of renewables***	-0.462
liquefaction	0.26	***transient modelling***	-0.462
power sectors	0.259	***life-cycle***	-0.463
wairakei	0.259	***pump system***	-0.463

fuel cycle	0.258	*u.s. energy demand*	-0.464
energy services	0.257	*hydrogen systems*	-0.465
storage medium	0.257	*modelling study*	-0.465
oil crisis	0.257	*correlation*	-0.466
energy conservation efforts	0.257	*business plans*	-0.466
energy conversion and management	0.256	*wind energy potential*	-0.467
government subsidy	0.256	*life cycle assessment*	-0.468
optimization models	0.256	*deregulation*	-0.468
high cost	0.256	*optimization models*	-0.468
green certificates	0.255	*coal combustion*	-0.469
energy shortages	0.255	*small communities*	-0.469
biomass energy	0.254	*electricity markets*	-0.47
electricity consumption	0.254	*electricity supply*	-0.47
watt-hours	0.253	*gas plants*	-0.472
fuel savings	0.253	*grid systems*	-0.473
pastoral properties	0.253	*lca*	-0.473
fossil fuel combustion	0.252	*drilling operations*	-0.473
fault detection	0.252	*natural ecosystems*	-0.473
low head	0.251	*sandstone*	-0.473
present situation	0.251	*conventional power*	-0.474
combined heat and power	0.25	*energy evaluation*	-0.474
hydroelectric dams	0.25	*delphi study*	-0.475
heating water	0.25	*wind resource*	-0.475
energy evaluation	0.25	*photovoltaic generation system*	-0.475
economic incentive	0.25	*residential buildings*	-0.475
conservation policies	0.25	*external energy*	-0.475
research efforts	0.25	*450*	-0.476
heat exchangers	0.25	*clean development mechanism (cdm)*	-0.476
state energy	0.249	*state energy*	-0.476
electricity markets	0.249	*scheikunde*	-0.476
renewable energy developments	0.247	*federal commitment*	-0.476
greenhouse gas abatement	0.247	*cross-flow*	-0.477
public perceptions	0.247	*general public*	-0.478
annual growth	0.247	*central australia*	-0.478
financial incentives	0.246	*policy options*	-0.479
gas-tight	0.246	*golden eagles*	-0.479
government policies	0.246	*power development*	-0.479
sound energy	0.246	*decision-making*	-0.48
bioresource technology	0.246	*efficiency measures*	-0.48
rural communities	0.246	*use of hydrogen*	-0.481
rotor diameter	0.246	*domestic energy*	-0.481
nongovernmental organizations	0.246	*service provider*	-0.482
technological aspects	0.245	*renewable energy usage*	-0.483
millennium development goals (mdg)	0.245	*sequencing batch*	-0.483
price of energy	0.244	*green certificates*	-0.483
distributed power	0.244	*rotor diameter*	-0.485
urban areas	0.244	*human health*	-0.485
wind energy conversion	0.244	*wave loading*	-0.486

project leaders	0.243	***financing processes***	-0.487
stream power	0.243	***integration system***	-0.487
developed countries	0.242	***geothermal drilling***	-0.487
technological learning	0.242	***united nations industrial development organization***	-0.489
local wind	0.242	***oil crisis***	-0.489
building designers	0.242	***petroleum reserves***	-0.489
energy conversion systems	0.241	***electricity use***	-0.49
advisory committee	0.241	***liquefaction***	-0.491
organisation for economic co-operation and development	0.241	***environmental impacts***	-0.491
reliable approach	0.24	***building energy efficiency***	-0.492
renewable energy (re) technologies	0.24	***production systems***	-0.492
photovoltaic array	0.24	***business reasons***	-0.492
built environments	0.239	***power production***	-0.492
electricity prices	0.238	***building stock***	-0.492
semi-structured interviews	0.237	***energy services***	-0.492
distributed generation	0.237	***renewable system***	-0.493
diffusion of renewables	0.236	***co-operative***	-0.494
spring water	0.236	***top-down***	-0.496
environmental benefits	0.236	***optimal placement***	-0.496
petroleum reserves	0.236	***energy conservation efforts***	-0.496
power system control	0.235	***impact energy***	-0.497
renewable energy program	0.235	***renewable energy literature***	-0.497
policy scenarios	0.235	***renewable energy penetration***	-0.497
short segment	0.235	***international conferences***	-0.498
higher educational institutions	0.235	***electricity demand***	-0.498
harmful emissions	0.235	***government subsidy***	-0.498
base load	0.234	***decision-support system***	-0.499
biogas system	0.234	***desalinated water***	-0.501
environmental conservation	0.234	***comparative analysis***	-0.501
sequencing batch	0.234	***fuel savings***	-0.501
electricity generator	0.234	***franchise tax***	-0.502
landfill gas	0.233	***distributed power***	-0.503
economic costs	0.233	***co-ordination***	-0.503
hydrothermal systems	0.233	***policy incentives***	-0.503
economic aspects	0.233	***power sectors***	-0.504
supply shortages	0.233	***international cooperation***	-0.504
agricultural sector	0.232	***uranium mining***	-0.505
gas reserves	0.232	***distributed energy generation***	-0.505
sensitivity analysis	0.231	***combined heat and power***	-0.506
gearbox	0.231	***employment growth***	-0.507
hydro-electric power	0.231	***hydroelectric dams***	-0.507
policy reform	0.231	***fuel cycle***	-0.508
planning phases	0.231	***negative effects***	-0.508
electricity production	0.23	***ecologically sustainable development***	-0.509
heat pump system	0.23	***tertiary education***	-0.509
potential resources	0.23	***energy intensive economy***	-0.509
support mechanisms	0.23	***nongovernmental organizations***	-0.509

conversion systems	0.23	*reservoir simulation*	-0.509
energy output	0.23	*stream power*	-0.509
renewable energy programmes	0.229	*power system control*	-0.51
siting	0.229	*small-scale renewable energy technologies*	-0.51
breeding populations	0.228	*outlays*	-0.51
decentralised development	0.228	*national economy*	-0.511
energy assessment methods	0.228	*solar engineering*	-0.511
energy performance regulations	0.228	*rural communities*	-0.512
environmentally friendly energy supply	0.228	*renewable electricity generation*	-0.512
golden eagles	0.228	*present situation*	-0.513
ministry of nonconventional energy sources	0.228	*renewable energy program*	-0.513
time-dependent	0.228	*watt-hours*	-0.515
financial assistance	0.227	*role of hydrogen*	-0.515
policy objectives	0.227	*hydro-electric power*	-0.515
generation systems	0.227	*spring water*	-0.516
social acceptance	0.227	*distributed generation*	-0.517
biogas plants	0.226	*conservation policies*	-0.518
intermittent	0.226	*power resource*	-0.519
rural livelihoods	0.226	*built environments*	-0.52
market participants	0.226	*landfill gas*	-0.52
rural areas	0.226	*energy shortages*	-0.52
energy requirement	0.225	*fossil fuel electricity generation*	-0.521
electricity sector	0.225	*government of india*	-0.521
environmental policies	0.224	*solar insolation*	-0.521
growth trends	0.224	*issue date*	-0.521
biomass resources	0.224	*spatial planning*	-0.522
local authorities	0.224	*economic incentive*	-0.522
economic co-operation	0.224	*questionnaire surveys*	-0.522
wave direction	0.223	*project leaders*	-0.523
power data	0.223	*financial incentives*	-0.523
electrolyzer	0.222	*harmful emissions*	-0.523
international agreements	0.222	*price of energy*	-0.523
finite energy resources	0.222	*energy conversion systems*	-0.523
hydroelectricity	0.222	*high cost*	-0.523
grid systems	0.222	*metal recovery*	-0.524
co-operative	0.222	*gearbox*	-0.524
role of hydrogen	0.221	*low head*	-0.524
geothermal power plant	0.221	*power system planning*	-0.525
commercial sector	0.221	*storage medium*	-0.526
non-governmental organizations	0.221	*electricity consumption*	-0.526
policy making	0.22	*poultry litter*	-0.527
performance prediction	0.22	*photovoltaic array*	-0.527
biomass energy sources	0.219	*fault detection*	-0.527
gathering information	0.219	*sound energy*	-0.527
collaborative initiatives	0.219	*performance prediction*	-0.528
replenish	0.219	*advisory committee*	-0.528
rural economy	0.218	*government policies*	-0.528

municipal utilities	0.217	*public perceptions*	-0.529
milieukunde	0.217	*renewable energy developments*	-0.529
environmental cost	0.217	*energy conversion and management*	-0.529
electricity industry	0.217	*research efforts*	-0.529
electric energy	0.216	*building designers*	-0.53
poultry litter	0.216	*environmental benefits*	-0.53
energy conversion efficiency	0.216	*urban areas*	-0.531
greenhouse issues	0.216	*conversion systems*	-0.532
significant wave height	0.216	*developed countries*	-0.532
renewable resource	0.214	*primary energy source*	-0.532
distribution networks	0.214	*aboriginal communities*	-0.532
optimization model	0.214	*policy reform*	-0.532
metal recovery	0.214	*distributed applications*	-0.532
economic developments	0.213	*pastoral properties*	-0.533
environment issues	0.213	*ground loop*	-0.533
wind energy conversion systems	0.213	*fossil fuel combustion*	-0.534
electrification	0.213	*option to purchase*	-0.534
indian renewable energy development agency limited	0.213	*biogas plants*	-0.534
hydrogen infrastructure	0.212	*steam power*	-0.534
long-term stability	0.212	*biomass energy sources*	-0.534
geothermal plant	0.212	*greenhouse gas abatement*	-0.534
economic performance	0.212	*geothermal power plant*	-0.534
promotion of renewable energy sources	0.211	*energy performance regulations*	-0.534
novel design	0.211	*support mechanisms*	-0.535
ground loop	0.21	*electric energy*	-0.535
solar engineering	0.21	*wind energy conversion*	-0.535
geothermal resources	0.21	*emission scenarios*	-0.535
experience curves	0.209	*wave energy conversion*	-0.535
ashrae	0.209	*heat pump system*	-0.536
hydroelectric energy	0.208	*local wind*	-0.538
united nations industrial development organization	0.208	*gas reserves*	-0.539
local processing	0.208	*generation systems*	-0.539
energy utilization	0.207	*rural areas*	-0.54
natural gas consumption	0.207	*wairakei*	-0.54
biomass gasification	0.207	*electricity prices*	-0.54
steam power	0.206	*hydrothermal systems*	-0.54
energy conservation measures	0.206	*wave direction*	-0.541
distributed applications	0.206	*electricity generator*	-0.541
form of energy	0.205	*short segment*	-0.542
geothermal fields	0.204	*waste energy resources*	-0.542
public acceptance	0.204	*environmentally friendly energy supply*	-0.542
producer gas	0.203	*municipal utilities*	-0.544
research priorities	0.203	*gas-tight*	-0.544
broad-winged hawk	0.202	*technological aspects*	-0.545
renewable energy literature	0.202	*electrification*	-0.545
industrial sector	0.202	*annual growth*	-0.546
wind energy potential	0.202	*geothermal resources*	-0.546

water demands	0.202	***ress***	-0.547
end-use efficiency	0.201	***supply shortages***	-0.548
role of technology	0.201	***local authorities***	-0.549
wind energy converter	0.2	***environmental policies***	-0.549
wood waste	0.2	***electricity production***	-0.55
cost reduction	0.2	***bioresource technology***	-0.55
fatigue life	0.2	***environmental cost***	-0.55
geothermal exploration	0.2	***siting***	-0.55
economic incentives	0.2	***indian renewable energy development agency limited***	-0.551
generation of electricity	0.199	***economic costs***	-0.552
government policy decisions	0.199	***battery consortium***	-0.553
total energy demand	0.198	***replenish***	-0.553
swine manure	0.198	***economic aspects***	-0.553
fusion energy	0.198	***hydroelectricity***	-0.554
farm operation	0.197	***space heating***	-0.555
energy impacts	0.197	***financial assistance***	-0.555
reactive power	0.196	***biomass resources***	-0.555
competing technologies	0.196	***base load***	-0.555
co-ordination	0.196	***form of energy***	-0.556
gas-fired	0.196	***wind energy conversion systems***	-0.556
electricity supply industry	0.195	***agricultural sector***	-0.557
environmental superiority	0.195	***distribution networks***	-0.561
geothermal gradient	0.195	***electrolyzer***	-0.561
enthalpy	0.195	***hydroelectric energy***	-0.561
nuclear science and engineering	0.195	***renewable resource***	-0.561
rated capacity	0.195	***rhyolite***	-0.562
system voltage	0.194	***energy requirement***	-0.564
price mechanism	0.194	***thermal water***	-0.565
electricity price	0.194	***geothermal park***	-0.565
development processes	0.193	***electricity sector***	-0.566
investment incentives	0.193	***energy output***	-0.566
renewable system	0.193	***higher educational institutions***	-0.567
biomass briquettes	0.192	***energy assessment methods***	-0.567
qualitative data	0.192	***hydrogen infrastructure***	-0.567
power produced	0.192	***natural gas consumption***	-0.567
fuel transportation	0.192	***planning phases***	-0.568
developing country	0.192	***reliable approach***	-0.568
thermal water	0.192	***social acceptance***	-0.568
battery consortium	0.192	***non-governmental organizations***	-0.569
remote areas	0.192	***intermittent***	-0.57
primary factors	0.191	***wood waste***	-0.57
renewable energy activities	0.191	***environment issues***	-0.57
installed power	0.19	***rural economy***	-0.57
methane yield	0.19	***potential resources***	-0.571
legal barriers	0.19	***geothermal plant***	-0.571
energy carrier	0.189	***policy making***	-0.572
geothermal wells	0.189	***policy scenarios***	-0.573

space heating	0.189	***time-dependent***	-0.573
multidisciplinary programs	0.188	***economic developments***	-0.573
long-term contracts	0.187	***producer gas***	-0.574
final energy	0.187	***energy utilization***	-0.575
awareness generation	0.186	***electricity industry***	-0.576
wind energy utilization	0.186	***biomass briquettes***	-0.576
solar lanterns	0.186	***market participants***	-0.576
conventional energy sources	0.185	***rotor speed***	-0.576
wind events	0.185	***technological learning***	-0.577
methane generation	0.185	***breeding populations***	-0.577
direct use	0.185	***rural livelihoods***	-0.577
pitch angle	0.184	***biogas***	-0.578
indian renewable energy development agency	0.184	***long-term stability***	-0.578
commercial risks	0.184	***drudgery reduction***	-0.578
rural electrification	0.184	***policy objectives***	-0.578
public benefits	0.183	***fall in price***	-0.579
environmental costs	0.183	***energy conversion efficiency***	-0.58
integrated resource	0.183	***commercial sector***	-0.581
market transformation	0.183	***fusion energy***	-0.582
geothermal water	0.183	***rural electrification***	-0.583
applied energy	0.183	***international agreements***	-0.583
performance indicators	0.183	***power data***	-0.583
exergy	0.182	***[jel] q48***	-0.584
load management	0.181	***sensitivity analysis***	-0.584
economic comparisons	0.181	***renewable energy activities***	-0.584
rinse water	0.181	***biomass gasification***	-0.584
reduction of carbon dioxide	0.181	***novel design***	-0.585
geothermal development	0.179	***gathering information***	-0.585
electricity suppliers	0.179	***renewable energy programmes***	-0.586
transmission networks	0.179	***rated capacity***	-0.586
horizontal axis	0.179	***semi-structured interviews***	-0.586
wind power density	0.178	***finite energy resources***	-0.587
electricity sales	0.178	***role of technology***	-0.588
institutional support	0.176	***electricity suppliers***	-0.588
financing costs	0.176	***gas-fired***	-0.588
resource potential	0.176	***fatigue life***	-0.588
solar energy materials and solar cells	0.175	***greenhouse issues***	-0.589
biomass combustion	0.175	***nuclear science and engineering***	-0.589
market barriers	0.173	***energy conservation measures***	-0.591
district heating	0.172	***ashrae***	-0.591
wind plants	0.172	***optimization model***	-0.591
international initiatives	0.17	***research priorities***	-0.592
power coefficient	0.17	***geothermal water***	-0.592
biomass potential	0.169	***decentralised development***	-0.593
renewable energy utilization	0.169	***indian renewable energy development agency***	-0.593
power system operation	0.168	***generation of electricity***	-0.594
exergy analysis	0.167	***economic comparisons***	-0.595

customer choice programs	0.167	***local processing***	-0.595
rhyolite	0.167	***economic performance***	-0.595
active power	0.167	***growth trends***	-0.596
legislative aspects	0.166	***desertification problem***	-0.597
exergy efficiency	0.166	***industrial sector***	-0.598
energy situation	0.166	***end-use efficiency***	-0.599
technical potential	0.166	***geothermal wells***	-0.599
small island developing states	0.165	***electricity supply industry***	-0.6
turbine generator	0.165	***millennium development goals (mdg)***	-0.602
subcontractor	0.164	***cost reduction***	-0.602
environmental interest	0.164	***significant wave height***	-0.603
organizational aspects	0.164	***public acceptance***	-0.603
biomass fuels	0.164	***government policy decisions***	-0.604
nameplate	0.164	***price mechanism***	-0.607
liquid column	0.163	***geothermal development***	-0.608
financial viability	0.162	***economic incentives***	-0.609
wind velocity	0.162	***developing country***	-0.61
load data	0.16	***geothermal exploration***	-0.611
funding mechanism	0.16	***remote areas***	-0.612
capital cost	0.159	***fuel transportation***	-0.612
progress in energy and combustion science	0.157	***long-term contracts***	-0.614
rural energy	0.157	***conventional energy sources***	-0.614
green pricing	0.157	***energy situation***	-0.614
rankine cycle	0.156	***investment incentives***	-0.616
renewable energy sources (res)	0.156	***power produced***	-0.616
rotor speed	0.156	***energy impacts***	-0.617
steam production	0.154	***reactive power***	-0.617
doe program	0.153	***renewable energy policy options***	-0.62
generation cost	0.153	***market transformation***	-0.621
diesel generators	0.153	***water demands***	-0.621
australian cooperative research center for renewable energy	0.152	***reduction of carbon dioxide***	-0.622
bibliometric measures	0.152	***biogas system***	-0.622
economic topic	0.152	***energy carrier***	-0.622
energy demand variations	0.152	***district heating***	-0.622
fuel charge	0.152	***experience curves***	-0.623
geothermal park	0.152	***stingray tidal stream generator***	-0.623
renewable energy policy options	0.152	***electricity price***	-0.623
renewable energy subsidy policy	0.152	***system voltage***	-0.624
rural electrification in senegal	0.152	***biomass combustion***	-0.625
statute for upgrading industries	0.152	***final energy***	-0.625
stingray tidal current generator	0.152	***environmental costs***	-0.628
stingray tidal stream generator	0.152	***renewable energy (re) technologies***	-0.628
thermal spring	0.152	***pv power systems***	-0.628
wairakei field	0.152	***promotion of renewable energy sources***	-0.63
working fluid	0.152	***public benefits***	-0.63
domestic cooking	0.152	***restructured electricity market***	-0.63
renewable energy scenario	0.152	***applied energy***	-0.632

energy and buildings	0.151	*competing technologies*	-0.633
electric power systems research	0.151	*direct use*	-0.633
ministry of non-conventional energy sources	0.151	*grid-connected systems*	-0.634
mechanical power	0.15	*economic topic*	-0.635
low energy density	0.15	*geothermal fields*	-0.635
pollutant emissions	0.149	*total energy demand*	-0.635
private sector participation	0.147	*international initiatives*	-0.636
restructured electricity market	0.147	*farm operation*	-0.636
dry steam	0.147	*transmission networks*	-0.637
pv power systems	0.145	*development processes*	-0.637
micro hydropower	0.143	*integrated resource*	-0.638
market restructuring	0.142	*progress in energy and combustion science*	-0.638
wind atlas	0.141	*rinse water*	-0.638
greenhouse heating	0.14	*remote area power supplies*	-0.638
international bank for reconstruction and development	0.14	*active power*	-0.639
electricity industry restructuring	0.138	*legal barriers*	-0.639
grid-connected systems	0.138	*enthalpy*	-0.64
distribution of electricity	0.138	*green pricing*	-0.641
incentive mechanism	0.138	*load management*	-0.642
promotion strategies	0.134	*wind events*	-0.643
geothermal reservoirs	0.131	*qualitative data*	-0.643
binary cycle	0.131	*commercial risks*	-0.643
donor funding	0.13	*turbine generator*	-0.644
geothermal areas	0.13	*primary factors*	-0.644
air infiltration	0.129	*resource potential*	-0.644
heat supply system	0.129	*swine manure*	-0.644
remote area power supplies	0.129	*installed power*	-0.646
electric loads	0.128	*wind energy converter*	-0.646
energy intensities	0.127	*diesel generators*	-0.647
waste energy resources	0.127	*low energy density*	-0.648
financing support	0.126	*horizontal axis*	-0.652
load factor	0.125	*power system operation*	-0.653
rated power	0.125	*geothermal gradient*	-0.653
regional energy planning	0.125	*liquid column*	-0.654
institutional structures	0.123	*financing costs*	-0.654
energy surplus	0.123	*exergy*	-0.655
institutional barriers	0.123	*performance indicators*	-0.655
salton trough	0.121	*tidal elevation*	-0.657
technology choice	0.121	*electricity sales*	-0.657
geothermal field	0.121	*wind plants*	-0.657
stand-alone power systems	0.12	*renewable energy utilization*	-0.659
environmental pollutions	0.119	*milieukunde*	-0.659
test reactor	0.119	*stand-alone power systems*	-0.66
critical loads	0.118	*exergy analysis*	-0.66
warm springs	0.118	*rural energy*	-0.66
office of energy research	0.117	*biomass fuels*	-0.66

renewable energy technology (ret)	0.117	*wind power density*	-0.66
salton sea	0.116	*organizational aspects*	-0.661
low wind speed turbine	0.116	*market barriers*	-0.662
digester gas	0.114	*biomass potential*	-0.662
drudgery reduction	0.114	*methane generation*	-0.664
employment factors	0.114	*wind energy utilization*	-0.667
quota-based systems	0.114	*environmental interest*	-0.667
wake effects	0.114	*pitch angle*	-0.668
electric demands	0.114	*solar energy materials and solar cells*	-0.671
united nations development program	0.111	*environmental superiority*	-0.671
lithuanian energy sector	0.111	*rankine cycle*	-0.672
irrigation water pumping	0.11	*incentive mechanism*	-0.673
geothermal reservoir	0.108	*generation cost*	-0.674
incentive measures	0.108	*technical potential*	-0.675
peak powers	0.105	*institutional support*	-0.676
amorphous silicon solar cells	0.103	*customer choice programs*	-0.678
broad policy guidelines	0.101	*methane yield*	-0.678
progressive utilization	0.101	*small island developing states*	-0.68
role of particles	0.101	*micro hydropower*	-0.68
geothermal waters	0.098	*doe program*	-0.686
emissions characteristics	0.097	*subcontractor*	-0.686
republic of croatia	0.092	*pollutant emissions*	-0.688
national energy planning	0.092	*nameplate*	-0.691
thermal springs	0.092	*market restructuring*	-0.691
large scale dissemination	0.091	*capital cost*	-0.693
non fossil fuel obligation	0.091	*dry steam*	-0.693
geothermal fluids	0.089	*multidisciplinary programs*	-0.694
wind speed data	0.089	*mechanical power*	-0.695
licensing procedure	0.089	*thermal spring*	-0.697
competitive power markets	0.088	*funding mechanism*	-0.698
geothermal area	0.087	*wind velocity*	-0.699
geothermal steam	0.086	*steam production*	-0.7
community attitudes	0.081	*exergy efficiency*	-0.701
competitive electricity supply market	0.076	*power coefficient*	-0.701
german renewable energy source act	0.076	*private sector participation*	-0.701
kizildere	0.076	*international bank for reconstruction and development*	-0.701
opec effect	0.076	*financial viability*	-0.702
open loop control system	0.076	*australian cooperative research center for renewable energy*	-0.704
power grid interconnection	0.076	*legislative aspects*	-0.705
renewable energy development plan	0.076	*renewable energy sources (res)*	-0.708
renewable energy development status	0.076	*load data*	-0.708
time-dependent systems	0.076	*binary cycle*	-0.709
revenue certainty	0.07	*wind atlas*	-0.71
economical potential	0.068	*awareness generation*	-0.713
uk nffo	0.067	*distribution of electricity*	-0.715
rural energy services	0.064	*ministry of non-conventional energy sources*	-0.715

levelized energy costs	0.063	*promotion strategies*	-0.716
pressure decline	0.061	*role of particles*	-0.716
solar regime	0.061	*domestic cooking*	-0.717
peak energies	0.057	*energy and buildings*	-0.721
residential-commercial sector	0.053	*electric power systems research*	-0.722
optimal renewable energy model	0.051	*office of energy research*	-0.722
renewable energy quota system	0.051	*competitive electricity supply market*	-0.725
rural energy enterprises	0.051	*geothermal field*	-0.726
thermal waters	0.045	*working fluid*	-0.728
lighting end-uses	0.038	*greenhouse heating*	-0.731
subsidy measures	0.033	*geothermal areas*	-0.731
overview of fuel cell	0.03	*renewable energy subsidy policy*	-0.733
elsevier (co)	0.025	*renewable energy technology (ret)*	-0.737
renewable energy systems cost	0.022	*energy surplus*	-0.741
policies formulation	0.022	*warm springs*	-0.745
renewable energy distribution	0.005	*salton sea*	-0.746
renewable energy company limited	0.001	*fuel charge*	-0.746
greenhouse climate control	0	*donor funding*	-0.75
underwater power generation	0	*amorphous silicon solar cells*	-0.752
ministry of economic affairs of taiwan	0	*financing support*	-0.754
dynamic life cycle assessment	0	*quota-based systems*	-0.758
liquid dampers	0	*geothermal reservoirs*	-0.759
regulatory law framework	0	*salton trough*	-0.759
aggregated wind turbines	0	*digester gas*	-0.76
bureau of energy of ministry of economical affairs	0	*technology choice*	-0.761
conventional generation cost	0	*rated power*	-0.766
education and trainings	0	*air infiltration*	-0.766
eirev	0	*regional energy planning*	-0.768
emissions (asphalt)	0	*electric loads*	-0.769
national power installation capacity	0	*low wind speed turbine*	-0.772
negative secondary impacts	0	*load factor*	-0.773
policy design impacts	0	*bibliometric measures*	-0.773
postgraduate university courses	0	*statute for upgrading industries*	-0.773
rd and d	0	*wake effects*	-0.775
renewable energy equipment procurement	0	*institutional structures*	-0.776
renewable energy plan 2012	0	*test reactor*	-0.778
renewable energy systems allocation	0	*electricity industry restructuring*	-0.78
tlcd	0	*heat supply system*	-0.788
undergraduate university courses	0	*energy intensities*	-0.79
1.27e15 j	0	*renewable energy quota system*	-0.792
accelerated depreciation plan	0	*renewable energy scenario*	-0.793
ad 2020-2021	0	*institutional barriers*	-0.798
asean economic sector	0	*united nations development program*	-0.8
australian cooperative research centre for renewable energy research program	0	*critical loads*	-0.803
biogas electricity conversion	0	*environmental pollutions*	-0.805
carbon dioxide discharges	0	*electric demands*	-0.806

cascade of hydro power plants	0	*peak powers*	-0.807
centre of appropriate technology	0	*wairakei field*	-0.809
co2 emission control policy	0	*irrigation water pumping*	-0.812
coherent renewable energy policy	0	*geothermal reservoir*	-0.817
commercial energy scene	0	*non fossil fuel obligation*	-0.817
cost/efficiency ratio minimisation	0	*incentive measures*	-0.825
crude oil supplies reduction	0	*geothermal waters*	-0.825
decarbonised world	0	*power grid interconnection*	-0.836
decentralized power systems penetration	0	*republic of croatia*	-0.837
domestic renewable energy generators	0	*licensing procedure*	-0.842
domestic technology manufacturing	0	*time-dependent systems*	-0.843
ecologically optimized transport systems	0	*thermal springs*	-0.853
electric power policy frameworks	0	*geothermal steam*	-0.861
electrical power facility (epf)	0	*national energy planning*	-0.863
electrical power test	0	*open loop control system*	-0.865
emission control policy	0	*large scale dissemination*	-0.87
energy density utilization	0	*employment factors*	-0.871
energy supply security considerations	0	*competitive power markets*	-0.874
energy-price signals reformation	0	*economical potential*	-0.877
energy-saving technical progress	0	*geothermal area*	-0.878
environmental friendly sustainable development	0	*emissions characteristics*	-0.878
finance pre-feasibility studies	0	*renewable energy development plan*	-0.879
financing risk premiums reduction	0	*wind speed data*	-0.883
flexible photo-voltaic solar cell	0	*geothermal fluids*	-0.889
fossil fuel politics	0	*kizildere*	-0.895
fuel for thought strategies	0	*lithuanian energy sector*	-0.895
future energy supply networks	0	*renewable energy development status*	-0.912
geothermal resources -- environmental aspects -- oregon -- prineville region	0	*opec effect*	-0.912
global renewable energy movement	0	*levelized energy costs*	-0.922
high activity subjects	0	*progressive utilization*	-0.936
household energy situation	0	*residential-commercial sector*	-0.942
hybrid diesel systems	0	*community attitudes*	-0.946
indian national programme	0	*revenue certainty*	-0.97
intermittent renewable energy penetration	0	*rural energy services*	-0.979
intrinsic technology problem	0	*uk nffo*	-0.988
kirklees councils	0	*thermal waters*	-1.01
load powering	0	*pressure decline*	-1.036
local employments	0	*rural energy enterprises*	-1.039
low activity subjects	0	*peak energies*	-1.045
marginalized mountain population	0	*broad policy guidelines*	-1.066
market adoptions	0	*solar regime*	-1.104
multilevel search engines	0	*lighting end-uses*	-1.12
natural gas supplies reduction	0	*optimal renewable energy model*	-1.132
nonpooled electricity trading	0	*overview of fuel cell*	-1.162
optimal renewable energy mathematical (orem) model	0	*subsidy measures*	-1.18

optimal renewable energy mathematical model	0	*elsevier (co)*	-1.201
petroleum crude oil consumption	0	*renewable energy systems cost*	-1.204
policies and measures to enhance use of renewable energy sources	0	*solar heat pump combination*	-1.209
power plant financing process	0	*web-based renewable energy courses*	-1.238
project datums	0	*cascade of hydro power plants*	-1.279
public lands -- oregon -- prineville region -- management	0	*future energy supply networks*	-1.279
recursive linear planning model	0	*local employments*	-1.348
recursive linear program	0	*policies and measures to enhance use of renewable energy sources*	-1.348
regional power market competition	0	*solar pv utilization*	-1.348
regulated rural energy concessions	0	*policies formulation*	-1.369
renewable energy development strategies	0	*china-india cooperation*	-1.423
renewable energy engineering curriculum	0	*rural electrification in senegal*	-1.734
renewable energy engineering system	0	*energy demand variations*	-1.844
renewable energy generating equipments	0	*stingray tidal current generator*	-1.913
renewable energy generation cost	0	*german renewable energy source act*	-2.052
renewable energy policy harmonisation	0	*renewable energy distribution*	-2.404
renewable energy product purchase	0	*renewable energy company limited*	-2.766
renewable energy promotion strategies	0	*ministry of economic affairs of taiwan*	-3.366
renewable energy systems reliability	0	*underwater power generation*	-3.488
renewable energy trading experience	0	*greenhouse climate control*	-3.535
renewable facilites	0	*liquid dampers*	-3.655
ring-fenced markets	0	*bureau of energy of ministry of economical affairs*	-3.696
social acceptance variation	0	*conventional generation cost*	-3.696
socio economic optimal renewable energy model	0	*national power installation capacity*	-3.696
solar heat pump combination	0	*policy design impacts*	-3.696
solar pv utilization	0	*postgraduate university courses*	-3.696
sustained renewable energy deployment	0	*renewable energy equipment procurement*	-3.696
tax revenue method	0	*renewable energy systems allocation*	-3.696
tertiary renewable energy education	0	*dynamic life cycle assessment*	-3.758
tertiary trained engineers	0	*rd and d*	-3.765
tertiary trained policy makers	0	*education and trainings*	-3.835
tertiary trained scientists	0	*emissions (asphalt)*	-3.904
time 2020 year to 2021 year	0	*negative secondary impacts*	-3.904
tourist ventures	0	*renewable energy plan 2012*	-3.904
uniform green power market mode	0	*undergraduate university courses*	-3.944
web-based renewable energy courses	0	*aggregated wind turbines*	-3.973
wind excitation	0	*regulatory law framework*	-4.767
windfarm market	0	*eirev*	-4.975
workshop in a box program	0	*tlcd*	-4.975

A.7 Latent Semantic Analysis

LSA concepts:

Concept 1 of 37	
Eigenvalue	86559
-------	
global warming	0.59149
Articles	0.53163
energy technology	0.41600
wind energy	0.37611
carbon emissions	0.14663
environmentalism	0.11627
environmental engineering	0.10039
earth science	0.07057
Donate	0.02906
Biofuels	0.02595

Concept 2	
Eigenvalue	39996
-------	
wind energy	0.68584
Articles	-0.56990
global warming	0.20395
environmental engineering	-0.18486
carbon emissions	-0.16017
environmentalism	-0.15862
earth science	0.15195
Donate	0.15165
Biofuels	0.09604
energy technology	-0.08025

Concept 3	
Eigenvalue	29711
-------	
carbon emissions	0.49354
Articles	-0.46009
global warming	0.39632
Biofuels	0.34675
wind energy	-0.33847
environmental engineering	0.25746
environmentalism	0.22942
Biofuel	0.14365
Donate	0.05704
testing phase	0.04302

Concept 4	
Eigenvalue	25348
-------	
energy technology	-0.64540
biofuels	0.40326
articles	0.37028
global warming	0.22892
biofuel	0.20243
environmental engineering	-0.19089
petroleum resources	-0.16839
space technologies	-0.16772
donate	0.16275
environmentalism	-0.13946

Concept 5	
Eigenvalue	23308
-------	
biofuels	-0.67133
global warming	0.43221
donate	0.29221
biofuel	-0.26514
energy technology	-0.26326
wind energy	-0.22036
earth science	0.16795
book reviews	-0.12532
renewable energy research	-0.11510
articles	-0.09420

Concept 6	
Eigenvalue	15788
-------	
carbon emissions	-0.50139
wind energy	-0.39979
energy technology	0.36495
environmental engineering	-0.33807
global warming	0.32736
environmentalism	-0.22666
petroleum resources	0.22447
space technologies	0.21909
biofuels	0.15434

biofuel	0.13004

Concept 7	
Eigenvalue	13033

Donate	-0.79396
earth science	-0.39650
global warming	0.28625
energy technology	-0.19918
Biofuels	-0.15382
petroleum resources	-0.15346
space technologies	-0.14964
Articles	-0.08182
wind energy	0.06788
environmental engineering	-0.06056

Concept 8	
Eigenvalue	9365

Biofuel	-0.72881
carbon emissions	-0.31870
Biofuels	0.31370
book reviews	0.27965
environmental engineering	0.27859
renewable energy research	0.20534
earth science	0.09906
environmentalism	0.09185
donate	-0.08776
editorials	0.07666

Concept 9	
Eigenvalue	9115
-------	
environmental engineering	0.62995
carbon emissions	-0.53592
biofuel	0.42983
environmentalism	0.22101
petroleum resources	-0.11971
space technologies	-0.11564
earth science	0.09156
chemical safety	0.08941
biofuels	-0.07759
failure analysis	0.07632

Concept 10	
Eigenvalue	6149
-------	
earth science	-0.50230
petroleum resources	-0.45964
space technologies	-0.43887
donate	0.34262
energy technology	0.25692
book reviews	-0.23827
editorials	-0.16914
biofuels	0.13549
biofuel	-0.12528
environmental engineering	-0.10575

Concept 11	
Eigenvalue	5671
-------	
earth science	-0.68537
petroleum resources	0.35246
space technologies	0.33423
donate	0.30206
energy technology	-0.21323
renewable energy research	0.18924
book reviews	0.18087
environmental engineering	0.16408
wind systems	0.12167
wind energy	0.10241

Concept 12	
Eigenvalue	5319
-------	
estonia	-0.53576
book reviews	0.50468
renewable energy research	0.40376
biofuels	-0.23292
editorials	0.22445
petroleum resources	-0.20995
biofuel	0.20808
space technologies	-0.19574
energy technology	0.14794
wind systems	-0.12999

Concept 13	
Eigenvalue	5200

-------	
estonia	0.83787
book reviews	0.32368
renewable energy research	0.24135
wind systems	-0.16343
editorials	0.15030
biofuel	0.13599
biofuels	-0.13414
energy technology	0.11337
petroleum resources	-0.11149
space technologies	-0.10324

Concept 14	
Eigenvalue	3675
-------	
wind systems	0.86193
editorials	0.37791
renewable energy research	-0.15330
book reviews	0.11565
wind energy	-0.11353
environmentalism	0.10209
chemical safety	-0.08578
energy technology	0.08020
environmental engineering	-0.08004
carbon emissions	-0.06591

Concept 15	
Eigenvalue	3630
-------	
microhydro	0.64025
renewable energy incentives	0.63675
chemical safety	-0.24479
editorials	-0.17378
failure analysis	-0.17070
legal environment	-0.16339
renewable energy research	0.11510
environmentalism	0.06372
articles	-0.05679
book reviews	-0.05040

Concept 16	
Eigenvalue	3439
-------	
chemical safety	0.61808
failure analysis	0.41669
legal environment	0.39908
renewable energy research	0.22661
pv-compact	-0.21262
renewable energy incentives	0.19556
microhydro	0.19323
wind systems	0.19320
editorials	-0.18679
biofuel	-0.13960

Concept 17	
Eigenvalue	3191
-------	
editorials	-0.59147
renewable energy research	0.50406
wind systems	0.28598
environmentalism	0.28163
environmental engineering	-0.19646
microhydro	-0.19529
renewable energy incentives	-0.19231
earth science	0.17896
chemical safety	-0.13820
pv-compact	0.11303

Concept 18	
Eigenvalue	2828
-------	
environmentalism	0.78750
environmental engineering	-0.39598
editorials	0.22272
wind systems	-0.21738
renewable energy research	-0.14376
testing phase	0.13770
chemical safety	0.12854
carbon emissions	-0.11756
wind energy	0.09241
failure analysis	0.08817

Concept 19	
Eigenvalue	2003
-------	
pv-compact	-0.95616
chemical safety	-0.16832
biofuel	0.13096

environmentalism	0.10129
book reviews	-0.08009
legal environment	-0.05673
microhydro	-0.05273
failure analysis	-0.05192
biofuels	0.05058
carbon emissions	-0.04321

Concept 20	
Eigenvalue	1637
-------	
book reviews	0.63664
renewable energy research	-0.55371
editorials	-0.52502
environmental engineering	-0.05177
environmentalism	0.05032
pv-compact	-0.03964
wind systems	0.02826
legal environment	-0.02708
testing phase	0.02654
research output	0.02597

Concept 21	
Eigenvalue	1533
-------	
research output	0.72389
access key	0.68769
pv-compact	0.03643
renewable energy research	0.02039
editorials	0.01902
book reviews	-0.01695
environmentalism	-0.00951
petroleum resources	0.00868
legal environment	0.00843
failure analysis	0.00705

Concept 22	
Eigenvalue	1164
-------	
testing phase	0.97266
environmentalism	-0.17111
energy politics	-0.14771
chemical safety	-0.03003
failure analysis	0.02701
book reviews	-0.01476
biofuels	0.01388
legal environment	0.01286
beryllium disease	-0.01212
chronic beryllium disease	-0.01205

Concept 23	
Eigenvalue	1133
-------	
chemical safety	-0.56798
failure analysis	0.54943
legal environment	0.39304
beryllium disease	-0.29361
chronic beryllium disease	-0.29203
vibration control	0.20523
pv-compact	0.04975
biofuel	-0.04370
testing phase	-0.03907
energy politics	0.01824

Concept 24	
Eigenvalue	980
-------	
vibration control	0.97115
legal environment	-0.10772
chemical safety	0.10274
failure analysis	-0.10112
carbon emissions	-0.07787
environmentalism	0.07486
beryllium disease	0.07024
chronic beryllium disease	0.06978
pv-compact	-0.02676
testing phase	0.02676

Concept 25	
Eigenvalue	569
-------	
energy politics	0.98664
testing phase	0.13665
environmentalism	-0.07903
biofuels	-0.01971
failure analysis	-0.01397
beryllium disease	0.01265
pv-compact	0.01255

chronic beryllium disease	0.01254
legal environment	0.01195
renewable energy research	0.00977

Concept 26

Eigenvalue	447
---	
beryllium disease	0.54996
chronic beryllium disease	0.54614
failure analysis	0.48954
chemical safety	-0.27283
legal environment	-0.24866
offshore wind turbine	-0.14879
vibration control	-0.02653
space technologies	0.01107
wind systems	0.00798
editorials	0.00713

Concept 27

Eigenvalue	443
---	
offshore wind turbine	-0.98746
beryllium disease	-0.08906
chronic beryllium disease	-0.08840
failure analysis	-0.06386
chemical safety	0.04588
wind systems	0.04145
legal environment	0.02451
carbon cap-and-trade	-0.01247
renewable energy incentives	-0.00947
wind energy	0.00924

Concept 28

Eigenvalue	377
---	
legal environment	-0.73959
failure analysis	0.44897
beryllium disease	-0.31065
chronic beryllium disease	-0.30621
chemical safety	0.23383
renewable energy incentives	0.04604
microhydro	-0.04475
energy politics	0.02204
editorials	0.01850
offshore wind turbine	0.01842

Concept 29

Eigenvalue	359
---	
renewable energy incentives	-0.70381
microhydro	0.69937
carbon cap-and-trade	-0.10363
legal environment	-0.05095
failure analysis	0.02835
beryllium disease	-0.01741
chronic beryllium disease	-0.01721
pv-compact	-0.01634
chemical safety	0.01403
offshore wind turbine	0.01397

Concept 30

Eigenvalue	282
---	
carbon cap-and-trade	0.99362
renewable energy incentives	-0.07558
microhydro	0.07181
biofuel energy	-0.03024
space technologies	-0.01644
petroleum resources	0.01521
offshore wind turbine	-0.01099
vibration control	0.01057
carbon emissions	-0.00944
environmentalism	0.00753

Concept 31

Eigenvalue	249
---	
biofuel energy	0.99531
space technologies	-0.05400
petroleum resources	0.05101
earth climate change	0.03652
carbon cap-and-trade	0.02843
biofuel	-0.02580
pv-compact	0.01887
legal environment	0.01326
failure analysis	0.01300
access key	0.00893

Concept 32

Eigenvalue	235
---	
space technologies	0.71929
petroleum resources	-0.68795
biofuel energy	0.07424
research output	0.03859
access key	-0.03380
carbon cap-and-trade	0.02383
failure analysis	-0.01755
legal environment	0.01275
energy technology	-0.00617
renewable energy incentives	-0.00593

Concept 33

Eigenvalue	166
---	
access key	0.72416
research output	-0.68760
space technologies	0.03731
petroleum resources	-0.03613
biofuel energy	-0.00852
offshore wind turbine	0.00253
failure analysis	0.00246
chemical safety	-0.00170
environmentalism	-0.00127
beryllium disease	0.00115

Concept 34

Eigenvalue	85
---	
ignalina nuclear power plant	-0.99988
estonia	0.01308
biofuel energy	0.00555
energy politics	0.00453
legal environment	0.00226
wind energy	0.00222
wind systems	-0.00172
offshore wind turbine	-0.00110
testing phase	0.00103
energy technology	-0.00085

Concept 35 of 37

Eigenvalue	76
---	
earth climate change	0.99925
biofuel energy	-0.03686
renewable energy research	-0.00681
book reviews	0.00517
editorials	-0.00379
environmentalism	-0.00339
energy politics	0.00296
earth science	-0.00253
testing phase	0.00230
environmental engineering	0.00218

Concept 36

Eigenvalue	2
---	
chronic beryllium disease	0.70993
beryllium disease	-0.70427
legal environment	0.00115
chemical safety	-0.00088
failure analysis	-0.00073
pv-compact	-0.00036
earth science	-0.00008
biofuel energy	-0.00008
environmentalism	0.00008
biofuel	0.00007

Concept 37

Eigenvalue	0
---	
energy-saving technical progress	0.99453
hybrid diesel systems	0.09332
centre of appropriate technology	0.04637
load powering	0.00578
renewable energy development strategies	0.00395
market adoptions	-0.00188
renewable energy generation cost	0.00072
biogas electricity conversion	0.00012
chronic beryllium disease	0.00000
beryllium disease	0.00000

References

[Aguilar, 1967] Aguilar, F. (1967). Scanning the business environment. New York: Macmillan.

[Anuradha et al., 2007] Anuradha, K., Urs, and Shalini (2007). Bibliometric indicators of Indian research collaboration patterns: A correspondence analysis. Scientometrics, 71(2):179–189.

[Berry et al., 1995] Berry, M. W., Dumais, S.T., and Letsche, T. A. (1995). Computational Methods for Intelligent Information Access. In Proceedings of the 1995 ACM/IEEE Supercomputing Conference.

[Cilibrasi and Vitanyi, 2006] Cilibrasi, R. and Vitanyi, P. (2006). Automatic extraction of meaning from the web. In IEEE International Symp. Information Theory.

[Cuhls and Sahlo, 2003] Cuhls, K. and Sahlo, A. (2003). Technology foresight – past and future, Journal of Forecasting, Vol. 22 (2003) pp.79-82

[Fahey and Narayanan, 1986] Fahey, L, and Narayanan, V.K (1986). Macroenvironmental analysis for strategic management. St. Paul, MN: West.

[Glenn, 1972] Glenn, J. C. (1972). Futurizing Teaching vs Futures Course, Social Science Record, Syracuse University, Volume IX, No. 3 Spring 1972.

[Kaya Firat et al., 2008] Kaya Firat, A., Woon, W.L. and Madnick, S. (2008). Technological Forecasting – A Review, MIT Sloan School Composite Information Systems Laboratory Working Paper #2008-15, September 2008.

[Landauer et al., 1998] Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An Introduction to Latent Semantic Analysis.

[Levary and Han, 1995] Levary, R.R. and Han, D. (1995). Choosing a technological forecasting method. Industrial Management, 37(1), 14.

[Morrison, 1992] Morrison, J. L. (1992). Environmental scanning. In M. A. Whitely, J. D. Porter, and R. H. Fenske (Eds.), A primer for new institutional researchers (pp. 86-99). Tallahassee, Florida: The Association for Institutional Research.

[Morrison et al., 1984] Morrison, J. L., Renfro, W. L., & Boucher, W. I. (1984). *Futures research and the strategic planning process: Implications for higher education* (ASHE-ERIC Higher Education Research Report No. 9). Washington, DC: Association for the Study of Higher Education. (ERIC Document Reproduction Service No. ED 259 692)

[Norton, 2001] Norton, M.J. (2001). Introductory Concepts in Information Science, New Jersey.

[Polanco, 1995] Polanco, X. (1995). Infométrie et ingénierie de la connaissance, in J.-M. Noyer (Ed.), Les sciences de l'information bibliométrie scientométrie infométrie, Rennes, Presses Universitaires de Rennes, 1995.

[Porter, 2005] Porter, A. (2005). Tech mining. Competitive Intelligence Magazine, 8(1):30–36.

[Porter et al., 2004] Technology Futures Analysis Methods Working Group, Technology Futures Analysis: Toward Integration of the Field and New Methods, Technology Forecasting and Social Change 71 (2004) (3), pp. 287–303.

[Rowe and Wright, 2001] Rowe, G. and Wright, G. (2001). Expert Opinions in Forecasting. Role of the Delphi Technique. In: Armstrong (Ed.): Principles of Forecasting: A Handbook of Researchers and Practitioners, Boston: Kluwer Academic Publishers.

[Saka and Igami, 2007] Saka, A. and Igami, M. (2007). Mapping modern science using co-citation analysis. In IV '07: Proceedings of the 11th International Conference Information Visualization, pages 453–458, Washington, DC, USA. IEEE Computer Society.

[Schoemaker, 1998] Schoemaker, P.J.H. (1998). Twenty Common Pitfalls in Scenario Planning, in Learning from the Future. Wiley & Sons, pp 422-431.

[Small, 2006] Small, H. (2006). Tracking and predicting growth areas in science. Scientometrics, 68(3):595–610.

[Woon and Madnick, 2008] Woon, W.L. and Madnick, S. (2008). Semantic distances for technology visualization, MIT Sloan School Composite Information Systems Laboratory Working Paper #2008-04, August 2008.

[Zhu and Porter, 2002] Zhu, D. and Porter, A. (2002). Automated extraction and visualization of information for technological intelligence and forecasting. Technological Forecasting and Social Change, 69(5).