

Global Comparison Aggregation Services (Web2002)

Hongwei Zhu
Stuart E. Madnick
Michael Siegel

Working Paper CISL# 2002-19

Novber 2002

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

Global Comparison Aggregation Services

Hongwei Zhu, Stuart E. Madnick, Michael D. Siegel
MIT Sloan School of Management
30 Wadsworth Street
Cambridge, MA 02142
{mrzhu, smadnick, msiegel}@mit.edu

Abstract

Web aggregation has been available regionally for several years, but this service has not been offered globally. As an example, using multiple regional comparison aggregators, we analyze the global prices for a Sony camcorder, which differ by more than three times. We further explain that lack of global comparison aggregation services partially contribute to such huge price dispersion. We also discuss difficulties encountered in the manual integration of global web sources. Motivated by this example, we propose a context mediation architecture for global aggregation to address semantic disparities of global information sources. Global aggregation services can bring efficiency to the global market and can be useful for market research and other business uses.

Keywords

Web Aggregation, Context, Semantic Integration

1. Introduction

With its increasing connectivity and capability, the World Wide Web is becoming the platform for global e-business. The global connectivity of the Web has not been exploited fully by existing service oriented e-business applications. For example, most of today's shopbots still only offer regional comparison shopping services, where regional (as opposed to global) information sources are used [1]. Comparison shopbots are also known as comparison aggregators for their capability of transparently aggregating information from multiple web sources [2].

What if comparison aggregation service is offered on a global basis? Imagine for the moment you are from Sweden and interested in buying a pocket sized digital camcorder. After some research on the Web you decide to buy a SONY DCR-IP5, which records video in MPEG format for easy editing on computers and weighs only 12 ounces (i.e., 336 grams). So you launch your favorite comparison aggregator to find the best deals and it returns information as shown in Figure 1.

Produkt	Butik	Märke	Pris	Lev.tid	Fraktpris	Totalpris
DCR-IP5 Visa mer	 Butikinfo	SONY	kr 19,495	2-5 D	kr 115	kr 19,610
DCR-IP5 Visa mer	 Butikinfo	SONY	kr 19,900	3-7 D	kr 75	kr 19,975
Sony DCR-IP5 Visa mer	 Butikinfo	SONY	kr 17,983	1-3 D	kr 99	kr 18,082
Sony DCR-IP5 (E) Visa mer	 Butikinfo	SONY	kr 19,115	1-3 D	kr 99	kr 19,214
Digital videokamera DCR-IP5. Visa mer	OnOff Butikinfo	SONY	kr 21,994	4-5 D	kr 95	kr 22,089

Fig. 1. Prices for DCR-IP5 in Sweden

Among the five vendors, 18,082 Swedish Krona (SEK) is the lowest total price. Is this the best deal, or is there a substantially better deal, on a global basis? Without a global aggregator, this can only be found out manually even with help of other comparison aggregators available in other countries. Our manual exercise found one vendor in the U.S. who ships the product to worldwide destinations at a total price of \$1,099.99 (\$999.99 plus \$100 international shipping charge, which includes warranty valid anywhere in the world), as shown in Figure 2.

Product	Options	Weight (lbs.)	Price	Qty.	Amount
		0.0 lbs.	\$999.99	1	\$999.99

Digital Camcorder SONY DCR-IP5
Total Quantity: 1
Subtotal: \$999.99

Shipping Information
*First Name: [I]
*Last Name: Zhu
Company Name:
Telephone:

Shipping Address
*Address: 303 Main
*City: Stockholm
*US State (if country is USA): -- Select a US State -- *Required if Country is USA
*Province: Stockholm *Required if Country is NOT the USA
*Zip/Postal Code: *Required if Country is USA
*Country: Note: Overseas shipments will incur a \$100.00 charge. Sweden

Billing Information is the same as Shipping? Yes No

Checkout

Fig. 2. An Offer for DCR-IP5 from the U.S.

Between 18,082 SEK and \$1,099.99, where would you buy? A seemingly simple question once you figure out that 1 US dollar is about 10 SEK. The Swedish offer is

64% more expensive than the U.S. offer. However, is this a special case? Again, a global aggregator will be helpful to provide an answer.

In the next section, we present a case study on worldwide price dispersion for the Sony camcorder. We explain various reasons why such dispersion exists and argue how global comparison aggregation can help connect global information sources, thereby bring efficiency to the global market. In section 3, we discuss deficiencies of existing comparison aggregation and technological challenges to providing global aggregation services. In section 4, we propose a scalable architecture that is promising for those challenges. We conclude with our insights of global comparison services in section 5.

2. Case Study – Global Price Dispersion for Sony DCR-IP5

One of the expectations of the European Union (E.U.) is to have an efficient integrated market with small price differences among member countries. A recent survey in the E.U. [3] shows that in the fresh food market “high price countries are often two times more expensive than countries with minimum prices”; even in the consumer electronics market, one country could be over 50% more expensive than another for a particular product. Data for that study was collected by three consultants who sampled various products in different stores.

Since comparison aggregation is a great tool for collecting price information, it has been used in a number of price dispersion studies in the U.S. for products such as books, CDs, and consumer electronics. Inter-store price differences were found to be 25-40% [4-6]. Although price dispersion still exists among online stores, overall online prices are lower than physical stores; for books and CDs, online prices were found to be 9-16% lower [4]. We could only find one study on price dispersion in the global online market [7], which showed that a U.S. buyer could save 42% for a particular textbook by purchasing it from the U.K. instead of from the U.S.

As there have been few studies on global price dispersion of the online market, we conducted an empirical study on the SONY digital camcorder in section 1: MICROMV DCR-IP5, which was introduced into the consumer electronics market in early 2002. Market prices for such a new product are extremely volatile; we took a snapshot of global prices by collecting data within 24 hours between March 8 and 9, 2002.

We used a number of regional comparison aggregators to retrieve the prices for the product. These aggregators

include BizRate, mySimon, Dealtime, Shopper, PriceRunner, PriceGrabber, Kelkoo, and Kakaku. We report our analysis on the unique vendor/price basis within a country. That is, if multiple aggregators in a country report on the same vendor, we treat them as one observation if the prices are the same or within \$1 difference. If a vendor has its online and physical stores as two entities, we treat them as two different observations even though both may charge the same price. All prices are listing prices not including shipping charges.

2.1 Worldwide Price Dispersion

We collected 172 observations covering US, Japan, and nine European countries. Figure 3 shows the histogram of prices. It is obvious that prices are highly dispersed. Most prices are within the range of \$1000-2000 and they are nearly evenly distributed in this range. Prices outside this range exist at both ends.

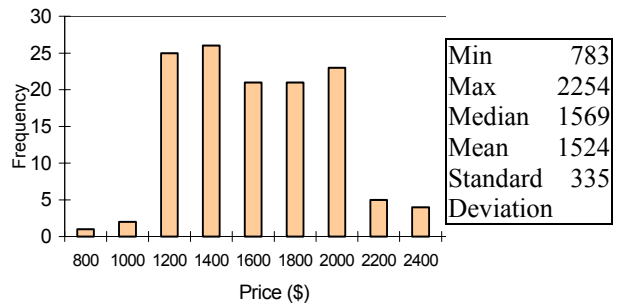


Fig. 3. Price Histogram (N=128)

Figure 4 shows the price distribution for all 13 countries, with the number of observations at the bottom. This is a box plot with each box representing 50% of price observations (i.e., the 25% and 75% quartiles) and the line within the box being the median. Lines stemming out of boxes cover all the other prices except for the extremes marked as solid circles.

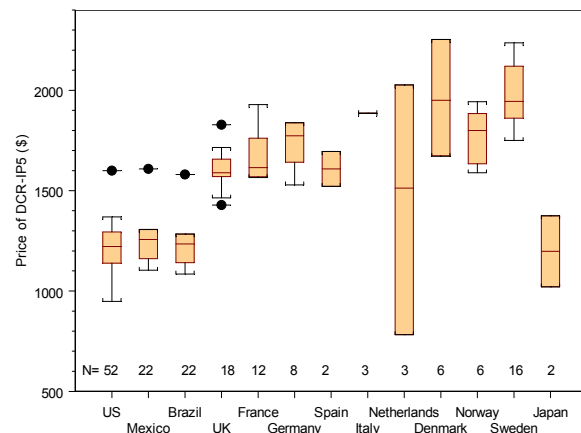


Fig. 4. Price distribution in different countries

Clearly, prices are different between countries. US and Japan have the lowest price levels. Most of continental European countries, except for Italy, have medium high prices. Italy and northern European countries have the highest price levels in our observation. Comparing with the international book price study [7], which shows that the U.K. has lower book prices, here we find that the U.K. has higher prices for this camcorder than the U.S.

Let's look at US prices in more detail, shown in Figure 5. These 53 unique price observations do not include SonyStyle US, Sony's online store in the U.S., and major consumer electronics vendors like BestBuy and CircuitCity, which offer the product at the same "official" price: \$1299.99. We can see from the figure most prices are at or below this price level. The average price is \$1203, which is 7.7% below the "official" price. More importantly, U.S. average price is 26.3% lower than the worldwide average.

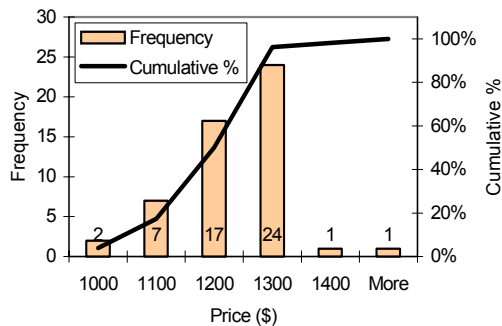


Fig. 5. Price Distribution in the U.S. (N=52)

2.2 Explanations for Price Dispersion

Textbook economic theory predicts that under perfect competition (e.g., Bertrand competition) commodity prices converge to one price, the so-called Law of One Price. But real world markets have shown no evidence to support this. The price dispersion phenomenon has been explained as a violation of one of the Bertrand assumptions: product homogeneity, zero search costs, or perfectly informed consumers [4].

In our case study, we looked at prices for one single product. Although it does have two models for video output (i.e., PAL and NTSC), this distinction is marginally important because its MPEG recording format allows for easy processing on a PC, which does not use the video output. In addition, many TV sets support dual video standards. So this product can be regarded as homogeneous. Regional aggregators can help lower search costs, which should lead to convergence of prices [8]. Whether all consumers are perfectly informed about price distribution is in question. Although comparison aggregation has gained some popularity, none of the popular comparison

aggregators ever make to the top 50 most visited sites in the U.S. measured by Jupiter Media Metrix.

In domestic e-business, it is possible that the three assumptions are met to some degree. In the context of global e-business, even the basic assumptions could be violated.

Although in terms of features the camcorder is nearly homogenous worldwide, other factors exist to result in heterogeneity. The product may be assembled in different plants that have different cost structures (e.g., plant in Malaysia vs. in Japan). Manufacturers often use different labeling to segment the market, e.g., different languages for product manuals in different regions. Warranty and other post sales services are often divided into regions.

Further, search costs are much higher due to lack of services that provide worldwide price information. We gave a hypocritical situation in the motivational example, but in reality chances are the Swedish buyer does not know any price information in the U.S.

These factors and the lack of a global comparison tool contribute to the worldwide price dispersion phenomenon. The following summarizes various explanations:

- Manufacturers have heterogeneous production costs around the world.
- Vendors have different pricing strategies, e.g., some may offer specials in certain parts of the world to promote sales.
- Buyers involve different search costs and have different preferences, e.g., buyers are not aware of price differences and weigh other factors more than price.
- Fluctuation of exchange rate causes price differences among countries.
- Manufacturer price control via market segmentation and other means of price discrimination, e.g., introducing product at different times.

Although price dispersion will not completely disappear, price transparency resulting from comparison aggregation should help mitigate dispersion and lower overall prices. This effect has been observed in the online market, e.g. average online prices are 7.7% lower than official price of the Sony camcorder and for books and CDs online prices are 9-16% lower than prices in physical stores [4]. Further, the U.S. average price for the camcorder is 26.3% lower than the worldwide average and the adoption rate of comparison aggregator in the U.S. is among the highest. Arguably, regional aggregation has helped increase competition and lower

the overall price level in the U.S. Global aggregation can potentially bring this efficiency to the global market, generating greater consumer benefits.

Next, we will examine the deficiencies of existing web aggregation services and identify technological challenges to advance from regional aggregation to global aggregation.

3. Technology Challenges to Global Aggregation

3.1 Deficiencies of Current Aggregation Services

Most of existing regional comparison aggregation is primarily implemented using web wrappers to extract information from web sources. This technology enables transparent aggregation even among non-cooperative sources, but conflicting implementation goals often deteriorate the quality of aggregated data. In addition, extraction tools do not address data semantics issues that are critical to service quality.

System responsiveness is often implemented at the price of compromising information timeliness, i.e., to achieve fast response to user request, many aggregators cache extracted data in their systems, resulting in out dated information. A daily update of the cache is not sufficient to avoid compromises on data timeliness because online prices change frequently due to low menu cost [9] and dynamic pricing strategies. Within a 2-hour window we observed a more than 2% decrease in average prices of the camcorder reported by one U.S. comparison aggregator [1]. Erroneous information can significantly impair the quality of comparison aggregation services. Figure 6 shows an example where a vendor has updated its price from \$62 to \$77.15 while the aggregator still reports the old price. If an automated purchasing agent decides to buy from the first vendor and makes the deal without consistency check, it will end up with overpaying 24%. This tension will become more outstanding in global aggregation because of increasing number of sources.

Store Name	Store Rating	Sort by Base price	+Approx Tax* & Shipping	Sort by =Total Price
powell's books Legendary Independent Bookstore	Be one of the 1st to review this store	\$62.00	Tax: check site Ship: \$2.50	check site
eCampUS.com EASY FAST CHEAP	★★★★☆ 4 Reviewer	\$70.45	Tax: check site Ship: €1.00	check site

powell's.com | HELP | YOUR ACCOUNT | CART | HOME

BROWSE SECTIONS | BESTSELLERS | SALE | USED | CITY CENTER

Rare & Collectible | Technical Books | Textbooks & Schooling | Kids

TECH BESTSELLERS | COMPUTER BOOKS | USED BOOKS

SEARCH

more search options

ABOUT THIS BOOK | Synopsis & Reviews

IN STOCK
Ships in 1 to 3 days
\$77.15

Fig. 6. Error due to Lack of Timeliness: Price Reported by Aggregator (top) vs. the Source (bottom)

Text based search, as opposed to semantics based search, can cause problems as well. In one occasion, an aggregator mistook a \$2 accessory of the camcorder and reported it as the price for the camcorder (although occasionally vendors make mistakes like \$1 laptops; but this is not the case here). Data semantics issues become more severe in global aggregation given the diverse contexts of sources, which we will explore next.

3.2 Data Semantics of Global Web Sources

The diversity in the origination and destination of information causes enormous complexity in making aggregated information meaningful and understandable. We have seen in the motivational example that currency needs to be converted using retail conversion rate to make sensible comparison for the Swedish user. Other issues exist. Let's illustrate these issues with an example of information about laptop computers from multiple sources of Sony, summarized in Table 1. We will ignore language difference in the following discussion.

Table 1. Information about a SONY Laptop Computer from Multiple Sources

	U.S.	U.K.	U.K. (in German)
Weight	2.76 lbs	1.26 kg	1,26 kg
Thickness	1.09"	NA	NA
Price	\$2,029 plus \$25 shipping	1,699.00 GBP incl. VAT	1.699,00 GBP inkl. MwSt.

First to note is that not all information is available at a single source. In this case the thickness information is not immediately available from U.K. sources (it is buried in a PDF document). If an aggregator takes the information from the U.S. source and directly reports to its German users, 1.09" probably would not be helpful to users who are familiar with metric systems for measurement. In addition to different units being used (lbs vs. kilograms, inches vs. millimeters, US dollars vs. British Pounds, etc.) there are other representational differences, such as symbols for thousands separator and decimal point. These differences have to be detected and reconciled for the users.

There is a more complicated problem in the data shown in Table 1. The last row shows pricing information for the product. Aside from representational differences, we note that the components going into price are quite different. Price, however simple as it appears, is in fact a complicated concept that has different meanings from different perspectives.

How much an item costs for someone to acquire is often different from how much it is listed for because of other costs that are associated with this transaction, including

taxes, duties if it involves international trade, shipping and handling, etc. An accurate calculation for price in the sense of “cost to acquire” could be very complicated in the context of global e-business. Calculation of VAT alone requires lots of additional information because VAT varies depending on the type of product, origination, destination, and special treaties between regions. The variations range from 0 to 25% of the listing price in European countries. The information listed in Table 1 is a hybrid of the two concepts for *price* with some missing components. This makes aggregation and meaningful comparison difficult. McCarthy and Buvac [10] illustrated this problem with an example of different prices of the same GE aircraft engine perceived by different organizations, such as the U.S. Air Force and U.S. Navy depending on whether the price includes spare parts, warranty, etc.

Another problem not explicitly shown in Table 1 is how the aggregators identify the same product from different regions. In the process of manually composing the Table, we noticed that the model numbers are different between laptops in the U.S. and those in Europe. We recognize their similarity (in this case identity except for the model numbers) by examining the configurations (e.g., CPU speed, hard disk capacity, weights, etc.). The fact that manufacturers often market the same product with different names in different regions makes it difficult for the aggregator to recognize their identity. This problem is best described from the following Camera example from Focuscamera.com:

“... a USA Minolta Maxxum is a Minolta Dynax overseas, the USA Canon EOS Rebel 2000 is an EOS 300 overseas, Pentax IQ Zooms are Pentax Espios overseas, etc.”

Conversely, when models with different features are named the same or slightly different in different regions, aggregators sometimes cannot recognize the distinction. In the Sony DCR-IP5 case study we found that some vendors label the product as DCR-IP5E to indicate that it is an international model compatible with the PAL standards rather than the NTSC standards in the U.S. What makes it worse is that most vendors use DCR-IP5 for both the NTSC model and the PAL model. Although this does not cause big problems because of its MPEG recording format, for other types of products this could be an issue.

The preceding discussions can be summarized into three categories of issues:

- Representation – how do we represent things
- Composition – what are the components for the thing

- Recognition – what is the thing we are really referring to

Next, we will propose an architecture that aims to address these issues so that users will get accurate, consistent, and meaningful aggregated information.

4. Architecture and Prototype for Global Aggregation Services

4.1 Context Mediation Architecture

The adoption of XML data standards and the emergence of Web services show promising signs for mitigating the tension between timeliness and responsiveness of global aggregation. But given the large scale and diversity of global aggregation, we recognize that heterogeneity of sources will continue to exist.

We propose a context mediation architecture (see Figure 7), which is based on the theories and techniques of context [10], mediated architecture [11], and the Context Interchange (COIN) project [12,13].

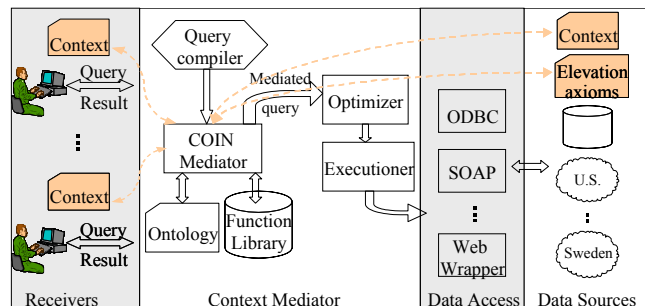


Fig. 7. A Context Mediation Architecture for Global Aggregation

Each online vendor or a regional aggregator is a data source, which can be accessed through the data access layer that implements various mechanisms to accommodate source heterogeneity. Both sources and receivers (i.e., users) have their contexts, which should be captured in logically distributed context knowledge bases. A common ontology or a set of aligned ontologies can be created by the aggregator. The mapping between data elements and ontologies is provided by elevation axioms. Contexts, ontology, and elevation axioms together address those three types of semantics issues. Conversion functions are used to translate values between contexts. The core part of a global aggregator is the COIN mediator, which resolves context conflicts between data sources and receivers.

With this architecture, a global aggregation user can specify what currency to use for price (representation) and whether the price includes or excludes taxes and

shipping handling (composition) about the specific product (recognition) offered by global vendors. Scalability is achieved by the abstraction of context and the modular design.

4.2 Prototype of Global Aggregation

A prototype has been developed to demonstrate the feasibility of the proposed architecture. We use a handful of regional aggregators as the sources. Using Cameleon web wrapper [14], we can impose a relational data model on these web sources and query them using SQL. For illustration purposes, our sources contain seller and price information for a single product – the SONY DCR-IP5 camcorder.

In the prototype, we focus on the issues of domestic and international taxation, shipping charges, and currency conversions that need to be addressed in global comparison services. Different situations of sources and receivers regarding these issues are represented as different contexts, samples of which are given in Table 2. They are axiomatized and recorded in the context knowledge base of the system. In addition, conversion functions are added to provide automated conversion service between contexts.

Table 2. Contexts for Price in Global Comparison

		Currency	Tax	Shipping ⁺
Source context	France	Euro	Included, 19.5%	Domestic: 15 Int'l: 80
	Sweden	Krona	Included, 25%	Domestic: 20 Int'l: 800
	UK	Pound	Included, 17.5%	Domestic: 10 Int'l: 35
	US	USD	Not included	Domestic: 50 Int'l: 100
Receiver context	US, Base	USD	Exclude	Exclude
	US, Cost	USD	If domestic vendor, no tax; otherwise, add 3% import tax	Include domestic or int'l shipping accordingly
	Sweden, Cost	Krona	Include 25% tax regardless	int'l shipping accordingly

⁺: Assume vendors only distinguish between domestic and interchange shipping charges. This is being refined to use online shipping inquiry services to calculate shipping costs by supplying product's dimensions and weight.

A domain ontology, as shown in Figure 8, captures the common concepts (in rounded boxes) and their relationships pertaining to contexts illustrated earlier. For example, a “seller” is a specialization of an “organization”, which has a “location” of type “country”. A modifier is a special attribute whose value

is specified in the context knowledge base. For example, “price” has a “type” modifier to indicate if it is base price, price with tax included, or total cost in a particular context.

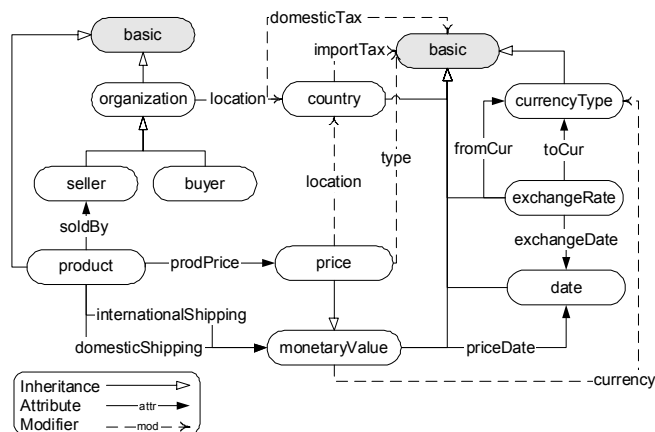


Fig. 8. Domain Ontology for Global Aggregator

A mapping between data from each source and the concepts in the ontology is provided by a set of elevation axioms to relate semantics to the data. All axioms and functions are supplied to a recent implantation of COIN mediation system [15], which can take user queries in SQL, automatically detect and reconcile context conflicts, and execute mediated queries to return results in the context of the user. We will give an example below to show how the system can help users such as the Swedish buyer mentioned in the beginning to do global comparison shopping.

The Swedish buyer is interested to know the total cost of the camcorder from worldwide vendors. His context has been recorded with the system. Now he can issue a query to compare prices of vendors all over the world using a predefined SQL, `compare_all`:

```
Select seller, price from kelkoofrance
union //French source
Select seller, price from pricerunnersweden
union //Swedish source
Select seller, price from pricerunneruk
union //UK source
Select seller, price from cnetshopper
union //US source
... //etc.
```

As we illustrated in the sample contexts, differences exist between the sources and the receiver. The COIN mediator automatically detects these differences and reconciles them by calling conversion functions. This process generates mediated queries that perform all the necessary conversions from source context to receiver

context. Some of the conversions we expect the system to automatically generate are given in Table 3.

Table 3. Appropriate Conversions for Reconciliation of Context Differences

Source	Conversion
France	Deduct 19.5% French tax, add 25% Swedish tax, add €80 international shipping, convert Euros to Krona
Sweden	Add 20 Krona domestic shipping
US	Add 25% Swedish tax, add \$100 international shipping, convert USD to Krona
UK	Deduct 17.5% UK tax, add 25% Swedish tax, add £35

The input SQL query is translated into a DATALOG query for the abductive reasoning engine to generate mediated queries in DATALOG, which in turn are translated into optimized SQL queries to be executed in parallel by the executioner [15]. The following gives the final mediated query automatically generated by the system to answer the user's initial query; we hope that readers can examine this and be convinced that all anticipated conversions are indeed performed by the following query. Note that `olsen` is an auxiliary online source that provides current and historical currency exchange rates; the system uses current date (*i.e.*, date when the query is issued).

```
//French source. Deduct 19.6% French tax; add 25% Swedish tax;
//add €80 int'l shipping; convert Euros to Krona
select kelkoofrance.seller,
(((kelkoofrance.price/1.196)+((kelkoofrance.pr
ice/1.196)*0.25))+80)*olsen.rate)
from (select seller, price
      from kelkoofrance) kelkoofrance,
//find exchange rate using auxiliary source
(select 'EUR','SEK',rate,'11/01/02' from olsen
 where exchanged='EUR'
 and expressed='SEK'
 and date='11/01/02') olsen
union

//Swedish source. Add 20 Krona domestic shipping
select pricerunnersweden.seller,
(pricerunnersweden.price+20)
from (select seller, price
      from pricerunnersweden) pricerunnersweden
union

//UK source. Deduct 17.5% UK tax; add 25% Swedish tax;
//add £35 int'l shipping; convert Pounds to Krona
select pricerunneruk.seller,
(((pricerunneruk.price/1.175)+((pricerunneruk.
price/1.175)*0.25))+35)*olsen.rate)
from (select seller, price
      from pricerunneruk) pricerunneruk,
//find exchange rate using auxiliary source
(select 'GBP','SEK',rate,'11/01/02' from olsen
 where exchanged='GBP'
 and expressed='SEK'
 and date='11/01/02') olsen
union
```

```
//US source. Add 25% Swedish tax; add $100 int'l shipping;
//convert USD to Krona
select cnetshopper.seller,
(((cnetshopper.price+(cnetshopper.price*0.25))+
100)*olsen.rate)
from (select seller, price
      from cnetshopper) cnetshopper,
//find exchange rate using auxiliary source
(select 'USD','SEK' rate '11/01/02' from olsen
 where exchanged='USD'
 and expressed='SEK'
 and date='11/01/02') olsen
union
...
```

An excerpt of the results is shown in Table 4 (reformatted from prototype output). All prices have been translated into the context of the Swedish user, who can easily compare them on the same basis. Finding the best deal globally is now as simple as clicking the predefined query with the help of this prototype of global comparison aggregation services.

Table 4. Excerpt of Results in User's Context

Source	Seller	Price (<i>i.e.</i> total cost in Krona)
Sweden	Foto & Elektronik AB	15815
	Expert Citybutiken/Konserthuset	16015

	Click ontime	23470
...
US	Bridgeviewphoto.com	10255
	PC-Video Online	10594

	Circuit City	14933

4.3 Extensions to Prototype and Related Issues

With this context-mediated architecture, a global aggregator can compare worldwide prices in a meaningful way for various users. This prototype successfully resolves representation and composition semantic conflicts. Recognition can be addressed by using a mapping of product codes to identify the exact product that may be labeled differently in various parts of the world. Alternatively, the system can use a formal ontology for products, which may become available on the Semantic Web in the future. Our future research on mediation using multiple ontologies and the development of the Semantic Web will help find alternative solutions.

The prototype can be readily extended to serve a broad audience by adding axioms for new sources and receivers. Clearly, technologies used here can enable full-scale implementation of global aggregation services, which will significantly increase the efficiencies of global e-business. Opportunities for

aggregation services are abundant. Readers interested in how aggregated information can be used to enhance values are referred to [16] for a thorough account. Data reuse plays an important role in the success of global aggregation services. The proposed COIN architecture provides solutions to technical challenges to reusing data from multiple web sources. Other obstacles still exist. Policy issues regarding data reuse are discussed in [17].

5. Conclusions

Despite the global presence of comparison aggregation, most of the services are offered regionally, not globally. Lack of global information can result in inefficiency in the global market. Our price dispersion case study shows that the worldwide prices for DCR-IP5, a Sony digital camcorder, can differ by nearly three times. A global aggregator can close the information gap and bring efficiency to the global market.

With this motivation, we propose a context mediation architecture to address data semantics issues for global aggregation. A prototype global aggregator has been developed to validate the architecture. The technologies used here show promising signs for building scalable platforms of global comparison aggregation services. These new services will benefit a variety of users. They will certainly help consumers find the best deals around the world; they can also assist researchers and policy makers to systematically collect market data with low cost (recall that the E.U. price dispersion survey mentioned in section 2 relied on three consultants who visited stores to manually collect retail prices); manufacturers can also use the services to find out the actual retail prices of their products around world, with which they can better assess demand and set appropriate wholesale and suggested retail prices. The emergence and the wide usage of global aggregation services will make the web the truly efficient platform for e-business.

Acknowledgement

The study has been supported, in part, by BSCH, Fleet Bank, Merrill Lynch, MITRE Corporation, the Singapore-MIT Alliance, and Suruga Bank

References

- [1] Zhu, H. (2002) "A Technology and Policy Analysis for Global E-Business", MIT Master's Thesis.
- [2] Madnick, S.E., Siegel, M.D. (2002) "Seizing the Opportunity: Exploiting Web Aggregation", *MISQ Executive*, 1(1), March 2002, 1-12.
- [3] EU Economic Reform (2001) "Price Dispersion in the Internal Market"
- [4] Brynjolfsson, E., Smith, M.D. (2000) "Frictionless Commerce? A comparison of the Internet and

- Conventional Retailers", *Management Science*, 46(4), 563-585.
- [5] Clay, K., Krishnan, R., Wolff, E. (2001) "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry", *Journal of Industrial Economics*, 49(4), 521-539.
- [6] Baye, M.R., Morgan, J., Scholten, P. (2001) "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site", Working Paper, Indiana University.
- [7] Clay, K., Tay, C.H. (2001) "Cross-Country Price Differentials in the Online Textbook Market", Working Paper, Carnegie Mellon University.
- [8] Bakos, J.Y. (1997) "Reducing Buyer Search Costs: Implications for Electronic Market Places", *Management Science*, 43(12), 1613-1630.
- [9] Bailey, J.P. (1998) Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce, Ph.D. dissertation, MIT.
- [10] McCarthy, J. and S. Buvac (1994) "Formalizing Context (Expanded Notes)", Stanford University.
- [11] Wiederhold, G. (1992). "Mediators in the Architecture of Future Information Systems." *Computer*, 25(3), 38-49.
- [12] Goh, C.H., Bressan, S., Madnick, S., Siegel, S. (1999) "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information", *ACM Transactions on Information Systems*, 17(3), 270-293
- [13] Madnick, S.E. (1999) "Metadata Jones and the Tower of Babel: The Challenge of Large-Scale Semantic Heterogeneity", *Proceedings of the 1999 IEEE Meta-Data Conference*, 1-13, April 6-7, 1999.
- [14] Firat, A., Madnick, S., Siegel, M. (2000) "The Cameleon Web Wrapper Engine", *Proceedings of the Workshop on Technologies for E-Services*, September 14-15, 2000, Cairo, Egypt.
- [15] Alatovic, T. (2002) "Capabilities Aware Planner/Optimizer/Executioner for Context Interchange Project", MIT Master's Thesis.
- [16] Madnick, S.E., Siegel, M.D. (2002) "Seizing the Opportunity: Exploiting Web Aggregation", *MISQ Executive*, 1(1), March 2002, 1-12.
- [17] Zhu, H., Madnick, S.E., Siegel, M.D. (2002) "The Interplay of Web Aggregation and Regulations", *Proceedings of 3rd International Conference on Law and Technology*, November 6-7, Cambridge, USA.