

Brokering infrastructure for minimum cost data procurement based on quality–quantity models[☆]

Alessandro Avenali^b, Carlo Batini^{c,*}, Paola Bertolazzi^a, Paolo Missier^d

^a *Istituto di Analisi dei Sistemi ed Informatica “A. Ruberti” — CNR, Italy*

^b *Dipartimento di Informatica e Sistemistica, Università di Roma “La Sapienza”, Roma, Italy*

^c *Dipartimento di Informatica, Sistemistica e Comunicazioni, Università di Milano Bicocca, Milano, Italy*

^d *School of Computer Science, The University of Manchester, UK*

Received 24 July 2006; received in revised form 9 October 2007; accepted 10 October 2007

Available online 4 November 2007

Abstract

Inter-organization business processes involve the exchange of structured data across information systems. We assume that data are exchanged under given condition of quality (offered or required) and prices. Data offer may include bundling schemes, whereby different types of data are offered together with a single associated price and quality. We describe a brokering algorithm for obtaining data from peers, by minimizing the overall cost under quality requirements constraints. The algorithm extends query processing techniques over multiple database schemas to automatically derive an integer linear programming problem that returns an optimal matching of data providers to data consumers under realistic economic cost models.

© 2008 Published by Elsevier B.V.

Keywords: Data quality; Information market; Information economics; Quality cost optimization; Bundle of data; Brokering service; Integer linear programming

1. Introduction

For large businesses and public sector agencies, good management of information assets has long been a key to their effectiveness in delivering quality services to users,

and many organizations have processes to manage the quality of their data. Recently, advances in the technology for large-scale deployment of information services, for example over service-oriented software infrastructures, have enabled cost-effective data exchange across organizations. In business terms, this means that it is becoming increasingly feasible for organizations to (i) purchase or otherwise acquire data from other peers, and (ii) exploit their own information assets for marketing purposes. These capabilities may be used to offer advanced services to users.

Thus, a general common trend is for organizations to acquire the information needed to support user services from third-parties. Several studies have analyzed the economic relevance of the potential information market.

[☆] The work presented in this paper has been partially supported by the eG4M MIUR FIRB project on e-Government in Mediterranean Countries and the MIUR FIRB MAIS project — Multi-channel Adaptive Information Systems: models, methodology, qualifying object-oriented platform and architectures for the flexible on-line information systems.

* Corresponding author.

E-mail addresses: avenali@dis.uniroma1.it (A. Avenali), batini@bicocca.mi.it (C. Batini), bertola@iasi.cnr.it (P. Bertolazzi), pmissier@cs.man.ac.uk (P. Missier).

Public agencies have been found to be the greatest producers of information by far, and the information they create and disseminate is often relevant for both the private and public processes, products, and services. In [33] an analysis of the commercial exploitation of *public sector information* is presented both for the USA and the European Union (EU). The study shows that the economic value of the information market in the EU for year 2000 amounted approximately to 10% of that of the US, where it was 750 billion dollars, and it recommended regulating the information market, to provide further incentives for the public sector information trading across and within member states.

To understand the implications of this trend, the size of the information market must be compounded with the issue of its *quality*, as a factor that will presumably affect the cost of data and hence the overall information market. Quality of data has been an issue since the nineties. General frameworks are available from the literature for describing data quality properties, or *dimensions* [37,38]. For instance, *accuracy* characterizes how well data represents its corresponding real-world entities. Another main issue concerned with information market is represented by offering *bundles* of data, which are indivisible units of data, each one with a single associated price and quality level. In fact, both the cost structure behind the production and the selling of digital information goods, and the necessity of implementing anti-competitive strategies can induce more and more data providers to offer indivisible units of different types of data (for example [28]). Focusing again on the public sector, it is well known that public agencies, in order to provide services to citizens and businesses, manage large registries with overlapping and heterogeneous data, and exchange large amounts of data flows. Such a huge number of registries, from one side is characterized by a high overlap, from the other side they are usually managed and updated with different policies, resulting in different levels of accuracy and other quality dimensions. In many data intensive processes sources are combined, and it is important for agencies and private users to be able to choose and compose data on the basis of the desired target quality. In other terms, the availability of such overlapping sources of data may be seen as an opportunity for the data demand, that may use a *quality driven query processing* strategy [27] that builds the global data set on the basis of the differentiated offer of data characterized by different qualities. Furthermore, the quality of data has a cost, and, at the same time, heavily influences the quality, the cost, and the revenues of the processes that use the data. While considering the relationship between the quality and cost of quality issues, some authors start their analysis from a

parallel between the emerging information market and established markets for other goods [7], with the final purpose of defining criteria for data quality control and improvement. These activities, like for other types of goods, have a cost which is a component of the selling price. Furthermore, in order to conceive rational methodologies for improving the quality of data, several authors have proposed data quality cost classifications [13,25] and *cost/quality optimization procedures* [6] that investigate the various different types of cost of non quality of data. Issues of quality driven query processing and cost/quality optimization have been addressed only recently so far. In particular, in the field of Decision Support Systems, the topic of data quality models to support and improve different situations of decision making has been studied in [36,10,20].

In this paper we propose a *brokering algorithm* that provides a *cost quality broker service* for facilitating the procurement of data from third parties, based on the assumptions that consumer interest for data is based both on its cost and on its quality, and that distinct data can be sold together in a bundle with a single associated quality and price. The algorithm, starting from: (i) the *offer of data* with possible bundling schemes from a set of providers, its quality and cost, (ii) the global, integrated knowledge on the information content offered by providers, and (iii) a query, that expresses the *data demand*, namely data requested by consumers and their quality, provides the optimal choice in terms of selected data, their quality and cost. We note that the broker service can be used as a decision support system for managers who have the responsibility of information acquisition activities.

The rest of the paper is organized as follows. In Section 2 the information procurement scenario underlying our approach is presented, together with a first overview of the algorithm is presented, and basic definitions. The two phases of the algorithm, decomposition and optimization, are detailed in Sections 3 and 4, respectively. A discussion on related work is presented in Section 5; Section 6 concludes the paper.

2. Overview of the approach and basic definitions

2.1. Information procurement scenario and cost model

To describe the input/output behaviour of the cost-quality broker, we introduce an *information procurement scenario* and a *cost model*. We assume the following:

- the information market consists of a potentially large number of organizations; each organization may have

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات