CLOUD COMPUTING AND BIG DATA AS CONVERGENT TECHNOLOGIES FOR RETAIL PRICING STRATEGIES OF SMEs

George SUCIU[*]
Gyorgy TODORAN[**]
Adelina OCHIAN[***]
Victor SUCIU[****]
Janna CROPOTOVA[*****]

Abstract

Most retailers know that technology has played an increasingly important role in helping retailers set prices. Online business decision systems are at the core point of an SMEs management and reporting activities. But, until recently, these efforts have been rooted in advances in computing technology, such as cloud computing and big data mining, rather than in newfound applications of scientific principles. In addition, in previous approaches big data mining solutions were implemented locally on private clouds and no SME could aggregate and analyze the information that consumers are exchanging with each other. Real science is a powerful, pervasive force in retail today, particularly so for addressing the complex challenge of retail pricing. Cloud Computing comes in to provide access to entirely new business capabilities through sharing resources and services and managing and assigning resources effectively. Done right, the application of scientific principles to the creation of a true price optimization strategy can lead to significant sales, margin, and profit lift for retailers. In this paper we describe a method to provide the mobile retail consumers with reviews, brand ratings and detailed product information at the point of sale. Furthermore, we present how we use Exalead CloudView platform to search for weak signals in big data by analyzing multimedia data (text, voice, picture, video) and mining online social networks. The analysis makes not only customer profiling possible, but also brand promotion in the form of coupons, discounts or upselling to generate more sales, thus providing the opportunity for retailer SMEs to connect directly to its customers in real time. The paper explains why retailers can no longer thrive without a science-based pricing system, defines and illustrates the right science-based approach, and calls out the key features and functionalities of leading science-based price optimization systems. In particular, given a cloud application, we propose to leverage trivial and non-trivial connections between different sensor signals and data from online social networks, in order to find patterns that are likely to provide innovative solutions to existing retail problems. The aggregation of such weak signals will provide evidence of connections between environment and consumer related behavior faster and better than trivial mining of sensor data. As a consequence, the software has a significant potential for matching environmental applications and business challenges that are related in non-obvious ways.

Keywords: *Cloud Computing, Big Data, Retail Pricing Strategies, Small and Medium Enterprise;*

---

[*] PhD candidate, Dipl. Eng, Faculty of Electronics, Telecommunications and Information Technology, 3CPS Department, University POLITEHNICA of Bucharest (e-mail: george@beia.ro).

[**] PhD candidate, Dipl. Eng, Faculty of Electronics, Telecommunications and Information Technology, TEF Department, University POLITEHNICA of Bucharest (e-mail: todoran.gyorgy@gmail.com).

[***] Dipl. Eng, Faculty of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest (e-mail: adelina@beia.ro)

[****] PhD Student, Dipl. Eng, BEIA Consult International, Bucharest (e-mail: victor.suciu@beia.ro).

[*****] PhD Student, Researcher, Practical Scientific Institute of Horticulture and Food Technology of Chisinau (jcropotova@gmail.com).

1. Introduction

The price setting process represents one of the key processes in a company and the aim of this process is to provide the mechanism to translate the longer term strategy of the company in terms of price positioning, market share goals, and product or service differentiation into the prices which are set or changed on a day-to-day basis within the competitive environment of the firm[1]. The key elements for short-term pricing decision making within the longer term strategic constraints of price positioning, volume goals, and product and service differentiation include costs, sales volume and its variation with the price, the impact of competitor price, and interaction between prices of certain products within the company's product portfolio.

Customer price knowledge has been the object of considerable research in the past decades[2]. Monroe and Lee[3] cite over sixteen previous studies, most of which focus on measuring customers' short-term price knowledge of consumer packaged goods. In a typical study, customers are interviewed either at the point-of-purchase or in their home and asked to recall the price of a product, or alternatively, to recall the price they last paid for an item. In perhaps the most frequently cited study, Dickson and Sawyer[4] asked supermarket shoppers to recall the price of an item shortly after they placed it into their shopping cart. Surprisingly, fewer than 50% of consumers accurately recall the price. Thus, despite the immediate recency of the purchase decision there is no improvement in the accuracy of the responses.

In another paper, Anderson, Cho, Harlam and Simester[5] combine survey data and a field experiment to investigate this prediction. In their study, they survey 14 customers and collect price recall measures for approximately two hundred products. They then conduct a field experiment in which they randomly assign the same items to one of three conditions. In the control condition, items are offered at the regular retail price. In the price cue condition, a shelf tag with the words "LOW prices" is used on an item. In the discount condition, the price is offered at a 12% discount from the regular price.

The authors show that both price cues and price discounts increase demand. But, consistent with theoretical predictions, the authors find that price cues are more effective on products for which customers have poor price knowledge. In contrast, price discounts are more effective when customers have better price knowledge. Together these results highlight the importance that price knowledge serves in determining the effectiveness of price changes and price cues.

2. Cloud Computing and Big Data Converging Technologies

The customers usually express their personal opinions regarding the products and services they purchase, this activity becoming a habit for many people nowadays. The online communities' continuous development, such as websites, blogs and social networks facilitate the exchange of information for the benefit of a growing number of users, increasing the social ties.

---

[1] Cassaigne, N., & Singh, M. G. (2001). "Intelligent decision support for the pricing of products and services in competitive consumer markets." IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 31(1), 96–106. doi:10.1109/5326.923272.

[2] A. Apostu, "Price conditions and price calculation in nowadays retail systems. Price knowledge extracted from Big Data," in Emerging Markets Queries in Finance and Business, 2014, unpublished.

[3] K. B. Monroe and A. Y. Lee, "Remembering versus knowing: issues in buyers' processing of price information," Journal of the Academy of Marketing Science, vol. 27, no. 2, pp. 207-225, 1999.

[4] P. R. Dickson and A. G. Sawyer, "The price knowledge and search of supermarket shoppers," The Journal of Marketing, pp. 42-53, 1990.

[5] E. T. Anderson, E. K. Cho, B. A. Harlam and D. Simester, "Using Price Cues," MIT, Cambridge MA, 2007.

On a growing market, the companies are forced to follow closely the needs and requests of consumers, but also their suggestions, in order to develop a more clear vision regarding the quality of products and services. The customer satisfaction has become a significant factor which affects the size and market segment profitability. In response to this problem, companies are forced to regularly survey the satisfaction levels of customers, wanting to focus on feedback from them and effecting improvements in work practices and processes used in the company[6].

Influencing consumers depends not so much the brand (although in the case of this tradition is a very important), but depends heavily on user reviews, opinions left by them on various specialized websites. Life cycle of a product on the market is influenced by the information about this product coming from consumers. This information can be a process because in a first phase, after the product is launched user reviews will appear, and these are always changing. Following these opinions, the most trusted and secure analyses are selected.

The amount of information available online is very large and grows exponentially with the development of social networks and virtual communities, in this way much of consumer feedback and ratings provided can be found online. Investing in IT solutions for the analysis of these large data volumes can be an optimal method to develop business environment[7].

2.1. Cloud Computing

The term cloud computing has been established relatively recent, in order to capture a particular use of IT resources, both software and hardware. There have been proposed a couple of definitions of this term, but the most relevant is that cloud computing refers to the use of computing resources as a service, over a network.

The cloud computing characteristics correspond to technologies or concepts which have been discussed individually since quite some time: virtualization, elasticity, and utility computing[8]. Virtualization means that an additional component within the software stack isolates the resource user from its concrete particularity. Elasticity means that a resource is always available to the user and that the resource grows or decreases when more or less of the resource is needed by the user.

There exist three service models for cloud computing, as infrastructure, platform and software and also four development models, as private, shared, public and hybrid, that together define the ways to deliver cloud services. The definition is intended to serve as a means for comparing cloud services and development strategies and to provide a basis for discussions about what is cloud computing and how it is more convenient to be used.

2.2. Big Data

Big data refers to the process of collecting and processing of very large data sets and to the associated systems and algorithms used to analyze these massive data sets. Architectures for big data usually range across multiple machines and clusters, and they commonly consist of multiple special purpose sub-systems. Coupled with the knowledge discovery process, big data movement offers many unique opportunities for organizations to benefit (with respect to new insights, business optimizations, etc.). Due to the difficulty of

---

[6] . Suciu, V. A. Poenaru, C. G. Cernat, G. Todoran and T. L. Militaru, "ERP and e-business application deployment in open source distributed cloud systems," in The 11th International Conference on Informatics in Economy (IE 2012), Bucharest, 2012.

[7] R. Eckstein, Interactive search processes in complex work situations: a retrieval framework (Vol. 10), University of Bamberg Press, 2011.

[8] G. Suciu, O. Fratu, S. Halunga, C. G. Cernat, V. A. Poenaru and V. Suciu, "Cloud Consulting: ERP and Communication Application Integration in Open Source Cloud Systems," in 19th Telecommunications Forum - TELFOR 2011, IEEE, Belgrade, 2011.

analyzing such large datasets, big data presents unique systems engineering and architectural challenges.

Currently, on the market there are many solutions for the search and analysis of large volumes of information solutions[9] which are usually focused on semantic technologies for aggregating and collating both structured and unstructured data[10]. Besides the well-known Google, there have been developed and are under development solutions for Enterprise Business Applications[11] such as CRM (Customer Relationship Management), ERP (Enterprise Resource Planning) and BI (Business Intelligence) and web applications, such as applications B2B (Business to Business), B2C (Business to Customer), using data from various sources (databases, web content, user generated content, etc.)[12].

Ontopica, a software company that uses the advantages of semantic technologies and the ontology development, in order to organize volumes of unstructured data, has released Dito[13], a software that provides a platform for online participation and browser-based search. Knime[14] is a platform for performing statistical data analysis and data mining the data for trend analysis and prediction of potential outcomes. RapidMiner[15] combines machine learning, data mining and predictive analytics being used more in research, education, development of applications.

2.3. Resulting Architecture for Big Data and Cloud Convergence

Exalead CloudView[16] represents a search platform which offers wide access to information on the infrastructure level and it is used for SBA (Search Based Applications) for both online and enterprise level. The application combines several semantic technologies for the applications development, but also analysis technologies (qualitative and quantitative), used on the data presentation level, aiming to provide the right information to the user.

CloudView represents a tool that combines search technologies with Business Intelligence, and it is a platform for the Exalead search engine, which was designed to implement semantic processing and selective navigation for the data volumes from online environments, facilitating processes such as searching and analysis, and also enabling organizations to improve their knowledge and resources exploitation.

Basically, CloudView represents an instrument that allows the exploitation of huge data volumes, both structured and unstructured. There are developed several connectors for these data sources, making them available for their presentation in an intuitive search interface. There is also used a modular index that combines structures, terminologies and semantic technology platform formats, providing a continuous access and unceasingly use of resources, through server distribution techniques and data redundancy[17]. The manner in which Exalead CloudView processes and offers access to information is illustrated in Figure 1.

---

[9] W. Li, Y. Zhong, X. Wang and Y. Cao, "Resource virtualization and service selection in cloud logistics," Journal of Network and Computer Applications, vol.36, no.6, pp. 1696–1704, 2013.

[10] T. Seymour, D. Frantsvog and S. Kumar, "History of search engines," International Journal of Management & Information Systems (IJMIS), vol. 15, no. 4, pp. 47-58, 2011.

[11] S. C. Yeh, M. Y. Su, H. H. Chen and C. Y. Lin, "An Efficient and Secure Approach for a Cloud Collaborative Editing," of Network and Computer Applications, vol. 36, no. 6, p. 1632–1641, 2013.

[12] M. Balazinska, B. Howe and D. Suciu, "Data markets in the cloud: An opportunity for the database community," in Proc. of the VLDB Endowment, 2011.

[13] Ontopica GmbH, "Dito," [Online]. Available: http://www.ontopica.de. [Accessed March 2014].

[14] R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter and T. Meinl, "KNIME-the Konstanz information miner: version 2.0 and beyond," SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 26-31, 2009.

[15] F. Jungermann, "Information extraction with rapidminer," in Proceedings of the GSCL Symposium'Sprachtechnologie und eHumanities, 2009.

[16] 3DS - Dassault Systèmes, "Exalead - Information Intelligence," [Online]. Available: http://www.exalead.com. [Accessed March 2014].

[17] G. Grefenstette and L. Wilber, "Search-based Applications: At the Confluence of Search and Database Technologies," in Synthesis Lectures on Information Concepts, Retrieval, and Services 2.1, 2010, pp. 1-141.
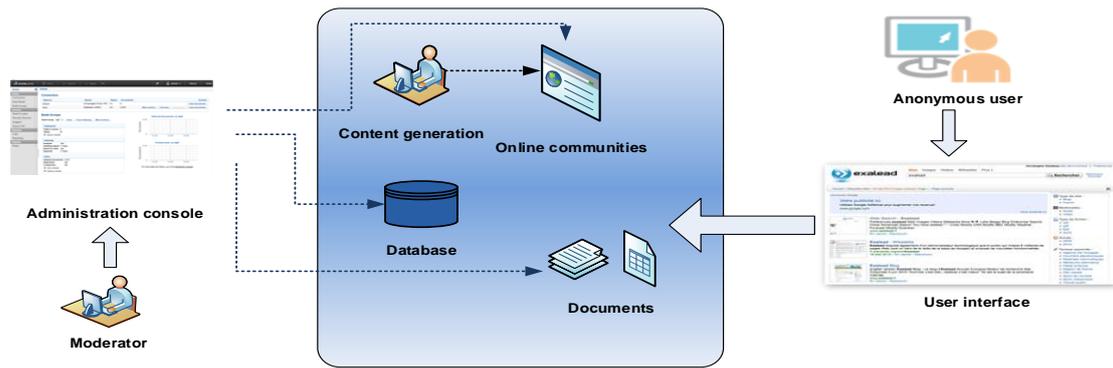
Figure 1. Processing and accessing big data via Exalead CloudView

CloudView accesses data from different sources using the so-called connectors. Each connector uses the data source protocol to connect its own information source and access the documents which will indexed. Through connectors, there can be indexed several data types by adding new documents, updating or deleting existing documents, extracting current list of indexed documents and managing data security.

CloudView provides multiple interfaces for data management and application configuration itself. These interfaces are represented by several APIs (Application Programming Interface), including a Push API for creating custom connectors, a Search API used for the development of other applications, Search and Management API for configuring and managing the indexing and search processes. Push API allows the indexing of any type of data, coming from any source, and supports basic operations required for the development of new connectors, both managed and unmanaged. A managed connector is part of the code that runs in CloudView, code that can be developed in Java. An unmanaged connector can be developed in any language, through CloudView's APIS, either Push APIs (in Java, C # and PHP) or through HTTP API.

Figure 2 provides a simplified view of the indexing process. The functionality can be described as follows: the connectors access the data sources and convert files into documents; these documents are sent to the Push Server and further are divided into groups, in the Analysis phase.
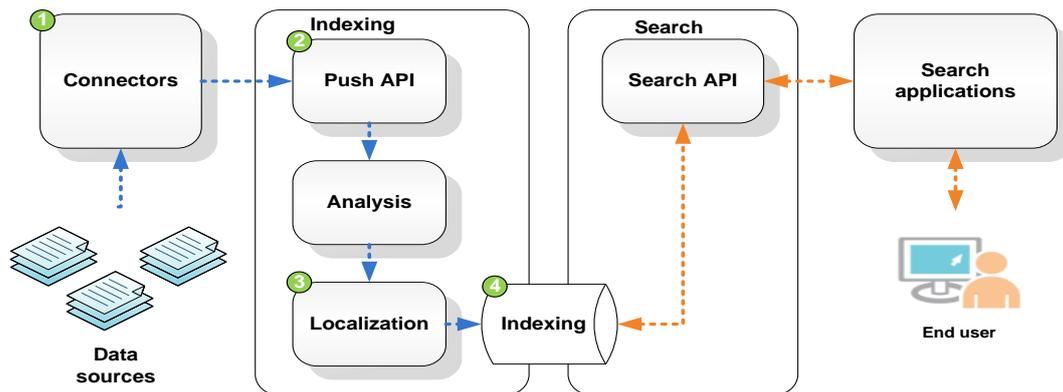

Figure 2. Simplified view of the indexing process

The analysis is performed sequentially, each document being processed, making retrieval of text, semantic processing, and other custom operations and location. The analysis also calculates the data to update the index. Once indexing is done and updated, the new documents are available for searching.

Connectors operate early in the CloudView indexing process, sending documents from different sources to a Push API server through the PAPI protocol. If the connector server is already existing, all new connectors will be automatically associated with it.

2.4. Configuration Methodology for CloudView

The CloudView configuration methodology is based on the connector's development, especially in the indexing process. This methodology is performed within the Administration Console, where the user can perform various configurations for the connectors, such as deleting all documents and restoring the connector's state, synchronizing indexed documents, or stopping the synchronization update for a specific type of connector (procedure that does not delete documents already processed). Standard connectors of CloudView include files, databases, HTTP, XML etc.

File type connectors are used for crawling local file systems or remote systems that are shared within the network. The search path configuration depends on the operating system on which CloudView runs In order to create a file type connector, it is necessary to ensure for the search system the access rights for the file system, i.e. file system rights to access the user account that is configured on CloudView. There also can be used regular expressions in order to specify a certain path (select *regexp*).

The JDBC database connector allows indexing of SQL database content, by using configurable queries for extracting data records. These records are then processed to compose the documents to be indexed. Documents are built by selecting the columns in the list returned by the synchronization queries. For indexing the database, a connector can be set to do a full synchronization, which indexes the entire database or set to one of the incremental modes for indexing only a part of the database. Incremental modes can be used when the table contains a timestamp representing the time of insertion or a unique serial number. When using incremental synchronizing, the connector takes the timestamp or the serial number from the column and assigns it to a variable in the query to retrieve new rows in the database.

HTTP connectors can load any URL address, provided that it is accessible from the server where CloudView is installed. The URL is taken, transmitted sequentially and then the links are extracted from that sequence. The main components of the CloudView system are the Crawler Server, Crawler Manager and Push Server. All other connectors, except those, that are hosted on the HTTP Server Connector, are installed on the Crawler Server. In the URL field, the site address is entered and some rules for indexing are set, for example reloading behaviour. Smart Refresh reloads URLs in the list and adds them to the queue, except for the recent uploads. URLs added by the reloading loop have a lower priority than newly discovered ones, so the reload does not slow the discovery of new pages.

2.5. The Data Model for the Price Optimization

The development of any search application in Exalead CloudView requires defining what data to include in the index schema, and then configuring one or more search logics to control how the documents are presented to users as search results. Index fields are used for searching and to display data in hit content in the results that are returned[18].

Categories store static facet values. These values display in the Refinements panel of the search results as well as in hit content. Static facets allow users to narrow their search results by focusing on a certain aspect of the results, such as a particular country or product line, price value within a range etc.[19]

[18] Shen Bin; Liu Yuan; Wang Xiaoyi, "Research on data mining models for the internet of things," Image Analysis and Signal Processing (IASP), 2010 International Conference on , vol., no., pp.127,132, 9-11 April 2010.
[19] Haesun Park; Van Huffel, S.; Elden, L., "Fast algorithms for exponential data modeling," Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on , vol.iv, no., pp.IV/25,IV/28 vol.4, 19-22 Apr 1994.

The data model represents a starting point for configuring the index and the search processes. Through a data model[20], it is allowed the management of different types of data according their purpose and their nature.

There can be created a set of property types like alphanumeric, numeric, geographic. After scanning the data sources, the settings which have been defined for the properties generate new index fields and categories inside the index schema and also several meta elements or prefix handlers according to the data logic.

When a configuration is applied, the high-level view provided by the Data Model is expanded into the multiple index and search elements. A common challenge when creating properties is to know which metas belong to the corpus. There are several ways in which the available metas can be explored using the CloudView data model.

When creating a data model property, it can be chosen one of the following field types: either an index field or a category facet only, or both. The most common indexing method is to assign a dedicatedIndexfield attribute in XML configuration to the property, represented in Figure 3. The property can then be made searchable (user queries can be applied to this field) and retrievable (the field can display in the search results).

Figure 3. XML configuration of the property

The property can also be stored as a facet and a category will be created with the value of the property. Faceted navigation will automatically be created for this property. In that case, it is possible to search for the value of the property. If the property belongs to the default class, the property can be searched with a prefix format of *property:value field*. Otherwise, it can be searched with a prefix format of *classname_property:value*.

---

[20] Shahriar, M.S.; Anam, S., "Quality Data for Data Mining and Data Mining for Quality Data: A Constraint Based Approach in XML," Future Generation Communication and Networking Symposia, 2008. FGCNS '08. Second International Conference on , vol.2, no., pp.46,49, 13-15 Dec. 2008.

The property can be made searchable without prefix, which means in addition to the indexing determined by other options, the property is indexed as searchable in the field called text. This allows users to search this property without specifying a prefix in the query.

When storing a numerical property as a facet, only equality search is possible. Range search is only possible in a dedicated index field. On numerical properties with a dedicated index field, searchable with prefix includes searching using the range operators.

While the data model simplifies the set up of the most common features for index fields and category facets, sometimes the user may need to customize indexing or search behavior. You can customize elements generated by data model expansion. For example, the user can modify the default clipping options for the data model-generated facets on a facet-by-facet basis.

3. Conclusions

In this paper we analyzed several existing solutions used for search and analysis of large volumes of data, with applicability in the retail field. Assessment of retail prices and product reviews can be developed using a search engine, by using semantic processing, which provides relevant results to the search query in case of incomplete or inaccurate results.

In order to create a model for the price evaluation, we developed several connectors and a data model based on properties and patterns which are likely to provide innovative solutions to existing problems. An evaluation can be based on classification opinions and calculation of notes, depending on the positive or negative reviews of their number, etc. Accessing this information can be made either online or by importing and compiling various formats (database, XML file).

References

- Cassaigne, N., & Singh, M. G. (2001). "Intelligent decision support for the pricing of products and services in competitive consumer markets." IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 31(1), 96–106. doi:10.1109/5326.923272
- A. Apostu, "Price conditions and price calculation in nowadays retail systems. Price knowledge extracted from Big Data," in Emerging Markets Queries in Finance and Business, 2014, unpublished.
- K. B. Monroe and A. Y. Lee, "Remembering versus knowing: issues in buyers' processing of price information," Journal of the Academy of Marketing Science, vol. 27, no. 2, pp. 207-225, 1999.
- P. R. Dickson and A. G. Sawyer, "The price knowledge and search of supermarket shoppers," The Journal of Marketing, pp. 42-53, 1990.
- E. T. Anderson, E. K. Cho, B. A. Harlam and D. Simester, "Using Price Cues," MIT, Cambridge MA, 2007.
- G. Suciu, V. A. Poenaru, C. G. Cernat, G. Todoran and T. L. Militaru, "ERP and e-business application deployment in open source distributed cloud systems," in The 11th International Conference on Informatics in Economy (IE 2012), Bucharest, 2012.
- R. Eckstein, Interactive search processes in complex work situations: a retrieval framework (Vol. 10), University of Bamberg Press, 2011.
- G. Suciu, O. Fratu, S. Halunga, C. G. Cernat, V. A. Poenaru and V. Suciu, "Cloud Consulting: ERP and Communication Application Integration in Open Source Cloud Systems," in 19th Telecommunications Forum - TELFOR 2011, IEEE, Belgrade, 2011.
- W. Li, Y. Zhong, X. Wang and Y. Cao, "Resource virtualization and service selection in cloud logistics," Journal of Network and Computer Applications, vol.36, no.6, pp. 1696–1704, 2013.
- T. Seymour, D. Frantsvog and S. Kumar, "History of search engines," International Journal of Management & Information Systems (IJMIS), vol. 15, no. 4, pp. 47-58, 2011.
- S. C. Yeh, M. Y. Su, H. H. Chen and C. Y. Lin, "An Efficient and Secure Approach for a Cloud Collaborative Editing," of Network and Computer Applications, vol. 36, no. 6, p. 1632–1641, 2013.
- M. Balazinska, B. Howe and D. Suciu, "Data markets in the cloud: An opportunity for the database community," in Proc. of the VLDB Endowment, 2011.

- Ontopica GmbH, "Dito," [Online]. Available: http://www.ontopica.de. [Accessed March 2014].
- M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter and T. Meinl, "KNIME-the Konstanz information miner: version 2.0 and beyond," SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 26-31, 2009.
- F. Jungermann, "Information extraction with rapidminer," in Proceedings of the GSCL Symposium'Sprachtechnologie und eHumanities, 2009.
- 3DS - Dassault Systèmes, "Exalead - Information Intelligence," [Online]. Available: http://www.exalead.com. [Accessed March 2014].
- G. Grefenstette and L. Wilber, "Search-based Applications: At the Confluence of Search and Database Technologies," in Synthesis Lectures on Information Concepts, Retrieval, and Services 2.1, 2010, pp. 1-141.
- Shen Bin; Liu Yuan; Wang Xiaoyi, "Research on data mining models for the internet of things," Image Analysis and Signal Processing (IASP), 2010 International Conference on , vol., no., pp.127,132, 9-11 April 2010
- Haesun Park; Van Huffel, S.; Elden, L., "Fast algorithms for exponential data modeling," Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on , vol.iv, no., pp.IV/25,IV/28 vol.4, 19-22 Apr 1994
- Shahriar, M.S.; Anam, S., "Quality Data for Data Mining and Data Mining for Quality Data: A Constraint Based Approach in XML," Future Generation Communication and Networking Symposia, 2008. FGCNS '08. Second International Conference on , vol.2, no., pp.46,49, 13-15 Dec. 2008