Bringing Innovative Semantic Technology to Practice: The iQser Approach and its Use Cases

Jörg Wurzer[1)], Bela Mutschler[2)]

[1)] iQser AG, Chlupfgasse 2, 8303 Bassersdorf, Switzerland
joerg.wurzer@iQser.net

[2)] University of Applied Sciences Ravensburg-Weingarten, Business Informatics
bela.mutschler@hs-weingarten.de

Abstract: This paper presents the iQser approach, a new semantic technology promising to overcome some of the shortcomings of current semantic enterprise solutions. The iQser technology is based on a powerful middleware platform (called iQser GIN platform). This practically-approved platform enables enterprises to efficiently develop effective semantic applications integrating various sources of structured and unstructured data. Goal is to enable users to gain new insights into complex information domains. In order to illustrate the practical relevance of the approach the paper also discusses potential enterprise computing use cases such as business (process) intelligence and information retrieval and also describes some demo applications that have already been developed.

1 Introduction

In their recent study *The Diverse and Exploding Digital Universe* IDC predicts that the amount of available digital information will reach 1800 Exabyte (1800 billion gigabytes) in 2011 (assuming that the amount of data will grow by 60 percent each year). This flood of information will not only make it more difficult to find needed information, it will also increase the complexity of dealing with information in many enterprises. In particular, it becomes increasingly challenging for enterprises to deal with new information influencing their business model.

Considering the access of information, keyword-based search functionality is currently widely used in enterprises, e.g., in intranets and intranet applications. However, as the potential result data are growing the result sets become long and diffuse. At the same time, extensive manual evaluations and selections of relevant information is economically not feasible for most enterprises.

Current approaches influenced by concepts of the Semantic Web do also not solve this problem because they heavily rely on well structured data and ontologies. Moreover, existing approaches of the Semantic Web are also limited by the data that can be used. Finally, most semantic solutions also rely on a keyword-based search paradigm.

Besides, shorter product life cycles, personalized products and services are other significant challenges enterprises have to deal with. Customers become a part of product development and product support and marketing using Web 2.0 tools and services becomes omnipresent. Hence, the demand for flexibility, the goal of reducing complexity, and dynamic market environments crave for new solutions for integrating and organizing information. Static structures like directories in document management systems or relational data model in traditional applications do not allow meeting these needs. Picking up basic notions of the Semantic Web, ontologies, for example, as an abstract layer are not a solution as well because of the dynamic and differing domains of information. Finally, there is a growing importance of the internet as a resource of information and a place of services for worth-drawing activities. Services and data are spread over different sources in the web. New concepts are required for a dynamic integration into business processes and enterprise applications. Meta search is not a solution because only the relations between integrated data and relations to business processes allow for the transfer of information into usable knowledge.

This paper presents the iQser approach, a new semantic technology promising to overcome some of the shortcomings of current semantic enterprise solutions. The iQser technology is based on a middleware platform (called GIN platform). This practically-approved platform enables enterprises to develop effective semantic applications integrating various sources of structured and unstructured data. Goal is to enable users to gain new insights into complex information domains.

The remainder of this paper is organized as follows. Section 2 presents the iQser approach in detail. Section 3 summarizes typical application scenarios the presented approach can support. Section 4 discusses the approach and explains why current semantic technology is scarce and why the iQser approach is needed in practice. Section 5 discusses related work. Section 6 concludes with a summary and an outlook.

2 The iQser Approach

For the iQser approach distributed data is brought together without any limitation of its format. The content of each information object is automatically analyzed without any invest in ontology modeling or semantic annotations. The main result of this analysis is an interconnectedness of the information objects to infer and provide implicit knowledge and deliver information in the action or working context of a user. This makes searching and collecting information from different sources unnecessary. In opposite of ontology-driven methods (top-down-approach) the iQser method derives structured knowledge from a data stock (bottom-up-approach). This is much more flexible and economical.

2.1 The GIN Platform

The GIN Platform is a proven semantic middleware as a basis for an efficient development of complex applications which use distributed data and their interconnections. It helps companies to save money for development and maintenance and users to save time in their daily working processes and get more out of their data.

2.1.1 Phase 1: Content Integration

The semantic middleware can integrate any data source like file-based data, webservices, databases and applications, which provide an API that can be accessed by Java. For each source the middleware needs at least one *ContentProvider* for a bidirectional integration of the data in this source. A *ContentProvider* is a kind of bridge that transforms the data into a generic information object for the middleware and vice versa, if that is part of the requirements. *ContentProviders* for some standard sources like common document formats are available out of the box.

In some cases a data source provides different types of information objects, for example the business objects in a CRM-system. There are several solutions for this fact. For each business object one could implement a *ContentProvider*, which is derived from a general *ContentProvider* for this source. In other cases like objects in a generic RDF-format only one *ContentProvider* could transfer and type all objects dynamically.

Unstructured or weak structured information objects are transformed into a structured information object by this process. Each object is semantically typed and has a list of attributes, which can be filled by metadata, text passages or data of media content like photos. For this process parsing and mining algorithms can be used additionally. A webpage in HTML for example can be parsed to get not only the tagged metadata but also structured information inside the body. For integrating articles of Wikipedia iQser identified for a demo application some more metadata like the correct title of the article or the categories.

For *ContentProvider* an unlimited set of actions or services can be defined. A basic action or service can be that a new or modified content object will be stored in a data base of the integrated source. That is useful for allow productive work with in an enterprise portal or to support processes where data has to be distributed in various systems.

By implementing a *ContentProvider* the integration of a data source is finished. There is no mapping with other data sources, no additional semantic annotation or ontology modeling needed. The bottom-up-approach is a kind of reengineering of the required data sources. This process can be used to design a master data model.

2.1.2 Phase 2: Data Consolidation

The GIN Platform provides a *Uniform Information Layer* as a central access point to retrieve and select information objects from all integrated sources. The selection of information objects can be specified by harmonized semantic types and metadata descriptors. If a user, for example, calls a list of customers, he can address this information objects by the semantic type „customer" independently from the definition in the data sources, which contain customer data and may have a different data model. Also the harmonized metadata can be used in this way.

Each information object is delivered in the same generic object format by the *Uniform Information Layer*. Therefore it is easy for a developer to build applications on top of the semantic middleware, because does not think about different data models, formats and services, which have to be identified and implemented. He just can define an application on the semantic level. After the middleware is installed in an organization and all data sources connected, new applications and process support can be realized in a short time with a small budget.

The *Uniform Information Layer* enables a meta search for users or any consumer system cross all integrated internal and external data sources including the web. For this information retrieval full text search is as well possible as search request with specified semantic types or metadata. Free combinations of this methods are either possible.

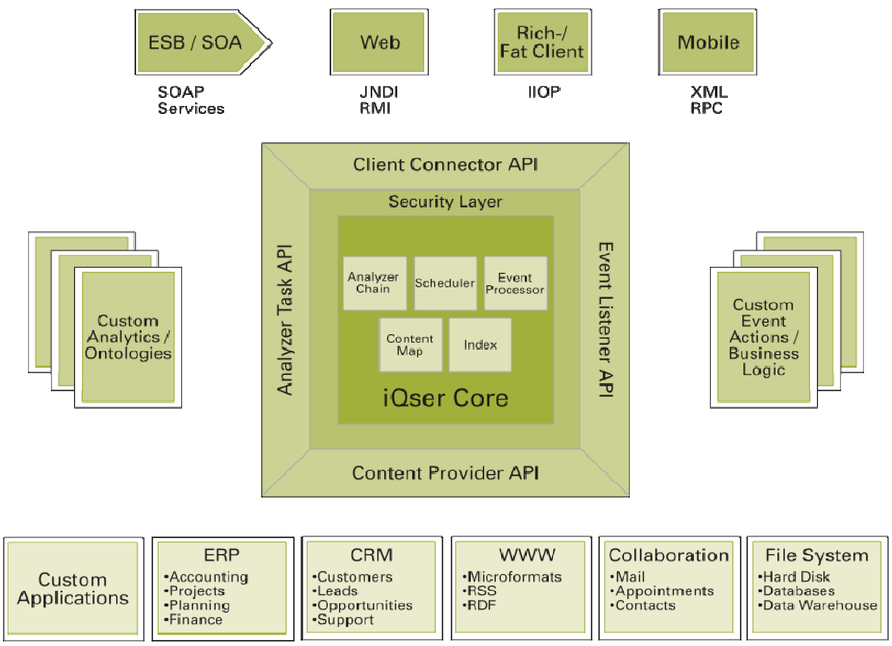

Fig. 1. The Architecture of the GIN Platform.

Figure 1 shows the architecture and the interfaces *Client Connector API* and *Content Provider API*, which have been discussed in the first two paragraphs 2.1.1 and 2.1.2. The *Client Connector API* enables access to the *Uniform Information Layer*. There are two more APIs shown in Figure 1: the *Analyzer Task API* and *Event Listener API*.

The *Analyzer Task API* offers developers to add some more analyzing methods to establish the object graph. This custom method can be used as a plug in for additional steps in the analyzer chain which is part of iQser Core. The *Event Listener API* is a useful interface to trigger or control a process that is defined by any business process engine. An event is fired if any new or changed data has been identified in the integrated data stock. So the middleware offers a central access to all events for a simple implementation of a designed process. *iQser Core* is the main part of the middleware. It contains the included analysis, indexing, event processing and calculation and persistence of the graphs. The analysis will be described in the next two paragraphs 2.1.3 and 2.1.4. Because of the full automated and totally enclosed features the developer need not take care about these things. The middleware makes it as easy as possible for the developers. A very important feature is the *security layer* of the GIN Platform. This considers all the access control management of an organization. The administrator need not define access rules again but the middleware uses existing access control management like Open Directory or LDAP. The user of an iQser-based application will get filtered information objects concerning his personal access rights.

A main difference to other integration platforms is the opportunity to call related objects of a selected information object. This feature enables a proactive, context sensitive information delivery and a more complex selection of information objects concerning their relations to infer implicit knowledge of the integrated data stack. Another main difference is the access to an abstract layer of the main concepts and their relations in the data stack. Both features will be explained in more detail in the next two paragraphs.

2.1.3 Phase 3: Deriving an Object-Graph

The object graph is the main result of the semantic analysis. He represents the complex relations of all information objects. Each connection between two information objects has a quantified relevancy and at least one reason with regards to the content. This additional information of the quadruple can be used to filter the connected information objects, for example for a focused view on the top five connected information objects in a given context. In other cases the additional information can be used to specify connections for a complex selection of information objects to query the implicit knowledge stored in the graph. It is a kind of reasoning on the graph by complex queries either describes in a graph structure.

The graph enables use cases beyond a time saving automation of information logistics. Here are some examples:

1. With the object graph the user can search in a given context. For example he looks for information, like documents which are related to a specific project. The starts a keyword search which is focused on results which are related to that project.

2. To search for the right ream members in a project staffing as project manager can identify people by their expertise which is not explicit described in their profile in an HR-system, but in indirect by the tracks they left in the company: participation in projects, published documents, problem solution, further education, etc.

3. A marketing manager can identify groups or classes of customers by statistical calculated common characteristics which are indirectly identified by connections to information objects like customer care requests, opinions in forums or (micro)blogs and so on.

The automized analysis has a high precision because of the combination of at three basic methods: metadata-matching, text text-similarity and usage-pattern. This method can be added by an open interface.

1. The metadata-matching only used metadata that is meaningful for connections between information objects. The middleware call them key attributes which are set in the *ContentProvider*.

2. The text-similarity should be named topic-similarity because the goal is to find texts with a similar topic focus. The method uses the fact that a topic is specified by the combination of significant concepts in a text. That solves the problem of semantic conflicts or synonyms.

3. The third method infers from the philosophy of language, that the meaning of language is created by the usage of language in a dialog. Therefore the third method analyses patterns of the usage of information objects to infer relations as well as a dynamic relevance of this relation.

The object adapts itself to the new and modified data stock and the changing information requirements of the users. Some information becomes currently more important and other information loses importance after a period of time.

2.1.4 Phase 4: Deriving a Concept-Graph

This analysis is separated from the analysis for object-graph by focusing on a conceptual level of all information and knowledge inside a data stock. Concepts comprise topics, classes and entities like names of people, companies, places, products and so on.

To build the concept graph the analysis calculates for each information object the most significant terms and their co-occurrences. The significance is calculated by the quotient of a time frequency in the analyzed object and the term frequency in the text corpus deduced from the index. After this step, the concept-graph is updated by the new values and the significance of co-occurrences is recalculated.

The result is a network of concepts with a quantified significance for each concept and co-occurrence. Therefore it is possible to calculate a hierarchical structure of the concept-graph such as a taxonomy with a free chosen concept as a root node or the most significant concepts without co-occurrences as a root layer.

This usage of the concept-graph can be used for a thematic overview by an automated taxonomy or semantic search by automated suggestions to specify the focus of interest. Further statistical reasoning on the concept graph are possible, for example the significance of a topic over a timer or new upcoming constellations of concepts.

3 Use Cases

This section sketches typical scenarios the iQser approach can effectively support in practice: (1) business (process) intelligence and (2) information retrieval.

3.1 Business (Process) Intelligence

Nowadays innovative products and services have to be developed under high cost pressure and time restrictions. This requires new types of dynamic collaboration scenarios within and between enterprises. Changes, either driven by internal or external factors, force business units to quickly adapt process-oriented information systems. In this context, the process-oriented alignment of information systems (e.g., enterprise resource planning systems) is success-critical. However, business processes are complex, rigid and adopting them to changed needs typically affects many people. Information systems are inflexible as well and very often implemented with the process logic "hard-wired" in the application code. Business (process) intelligence tools offer promising perspectives in this context [1].

Business intelligence (BI) refers to skills, technologies, applications and practices used to help a business acquire a better understanding of its commercial context. Business intelligence may also refer to the collected information itself. BI applications provide historical, current, and predictive views of business operations. Common functions of business intelligence applications are reporting, OLAP, analytics, data mining, business performance management, benchmarks, text mining, and predictive analytics. Business intelligence often aims to support better business decision-making.

More specifically, *business process intelligence* (BPI) applies business intelligence concepts (e.g., analytical applications) to business processes [2]. It is based on the analysis of process execution data and on the automatic derivation of (optimized) process models and process performance characteristics from these data. It is implemented as a set of integrated tools providing features for the analysis, mining, prediction, control, and optimization of processes. The overall goal of BPI is to extend an enterprise's performance management to business processes. Enterprises more and more realize that gaining knowledge about their processes results in many benefits, justifying the high costs arising from the introduction of BPI solutions. In fact, BPI tools offer several benefits. For example, business processes can be monitored during their execution, process optimization potentials can be derived in (nearly) real-time, and process information can be visualized in an aggregated form for different user groups (e.g., using dashboards). Examples of such tools include *Websphere Business Integration Monitor*, *ARIS Process Performance Manager* and *BizTalk Server Business Activity Monitoring Framework*. BPI tools utilize respective metrics to derive (aggregated) process information and to generate status reports. In the following we introduce three use cases for BPI and discuss the benefits arising in this context.

Use Case 1: Information System Alignment. BPI can be used to support the development and maintenance of process-oriented information systems. In particular, it provides valuable information (e.g., about the adequacy of provided business functions) for aligning the information systems to the business processes. The use of BPI tools can decrease software development costs.

Use Case 2: Business Process Optimization. **BPI can be used to identify "critical" s**cenarios that may occur during the execution of a business process (e.g., non-availability of resources, unnecessary waiting and idle times). Process mining as an important BPI concept allows for the continuous derivation of optimized process models. This, in turn, reduces the total effort necessary for "manual" process analyses. As optimizations are based on real data, their implementation tend to be much more effective than other approaches (e.g., the disclosure of optimization potentials by process simulation based on estimated data).

Use Case 3: Process Transparency/Visualization of Process Information. Due to the fragmented support of business processes, their control is distributed over several operational systems, i.e., we cannot always assure that controlled execution by one control system (e.g., process management system) is possible. Nevertheless, when collecting the respective log data from the different systems, it becomes possible to provide monitoring and visualization support for the overall business process. The information to be visualized include complete process schemas and process instances (e.g., control and data flows, activity states) as well as other process-related data (e.g., application data). Most BPI tools include features to visualize processes and related aspects. ARIS PPM, for example, offers a detailed tree view to illustrate the hierarchical relationships between processes and sub processes. Particularly the analysis of entire process maps becomes easier using such or comparable features.

However, contemporary visualization modules of BPI tools still lack some important issues: visualization of processes running on different platforms as well as features to customize process visualizations. Regarding the latter, for example, it is an objective to better adjust the process information to be visualized for different purposes. Process views, for example, target at providing either context-aware (e.g., visualization of control flow) or user-aware (e.g., high-level aggregated process data for managers) process information. Providing such advanced visualization features will further increase the advantages of this use case. In summary, the visualization and monitoring of distributed processes is a complex, though quite useful BPI use case.

As can be seen, BPI tools support a broad spectrum of use cases. However, the success of BPI tools in practice is not yet conceivable, though there is a growing interest of both vendors and customers for the topic. One promising approach is it to extend BPI tools with a powerful semantic technology such as iQser. This enables business analysts to gain new and better insights into analyzed business processes and their semantics.

Providing effective IT support for business processes has become crucial for enterprises to stay competitive in their market [3]. In the automotive domain, for example, a broad spectrum of business processes, ranging from simple administrative procedures to very complex, knowledge-intense engineering processes has to be effectively supported. Similar scenarios exist in many other domains like e-commerce, transportation, or healthcare. In all these cases, domain-specific processes must be defined, implemented, enacted, monitored, and continuously adapted to a changing context. Thus, process life cycle support and continuous process improvement adopt a key role in contemporary and future enterprise computing [4]. The process life cycle starts with the (re)design of a business process. Process modeling and process analysis tools can be used during this phase. Thereafter, the business process has to be implemented resulting in a process-oriented IS. As a typical example consider a *product data management* (PDM) system which offers a broad range of business functions to deploy models and documentation of the managed product(s) to involved user groups (e.g., engineers, managers, suppliers). Following the implementation and deployment phase, multiple instances of the implemented business process can be created and executed during the enactment phase. Finally, process enactment logs can be analyzed and mined to identify potentials for process optimizations. Using the iQser platform enables enterprises to better align their business processes to their process-aware information systems - both internally and in cross-organizational settings. For example, consider CRM systems. Using the iQser middleware, different data sources (CRM, document management, customer information from the internet, Lotus Notes) can be easily integrated and used for further semantic analysis (based on a derived concept graph).

3.2 Information Retrieval

Today, users are typically offered more or less fixed hierarchical navigation structures they can use to access information. Besides, keyword-based search functionality is offered as well. Often, users have many problems with this procedure.

Based on the iQser approach, it does not become necessary anymore to explicitly search for required information (e.g., using hierarchical navigation structures or keyword-based search functionality). The iQser middleware automatically delivers required information in a given context. The context may be characterized by assigned tasks, involved projects, customer requests or related documents.

As one typical example, consider the availability of process information in enterprises. It is a big challenge for many enterprises to provide the right process information at the right time for the right user. Problem is not only to identify the relevant information, but also to integrate the various sources of process information (e.g., textual process descriptions, official policies, graphical process diagrams, etc.). Applying the iQser middleware enables enterprises not only to flexibly integrate various sources of both structured and unstructured process information, but also to automatically provide business users with needed information without requiring to explicitly search for the respective information. As examples of respective information consider information on the next activities in a process, the roles assigned to process activities, or the documents are needed to execute a certain process activity.

A prototype application illustrating iQser-based information retrieval is shown in Figure 2. Specifically, this application deals with an automatic semantic analysis of wikipedia articles about astronomy. Goal is to demonstrate how knowledge can be extracted out of unstructured data and how this knowledge can be accessed in a new way by using the notion of *concept graphs* and *object graphs*. Both kinds of graphs (cf. Section 2) are the basis for reasoning, inferring, classification, and information delivery.

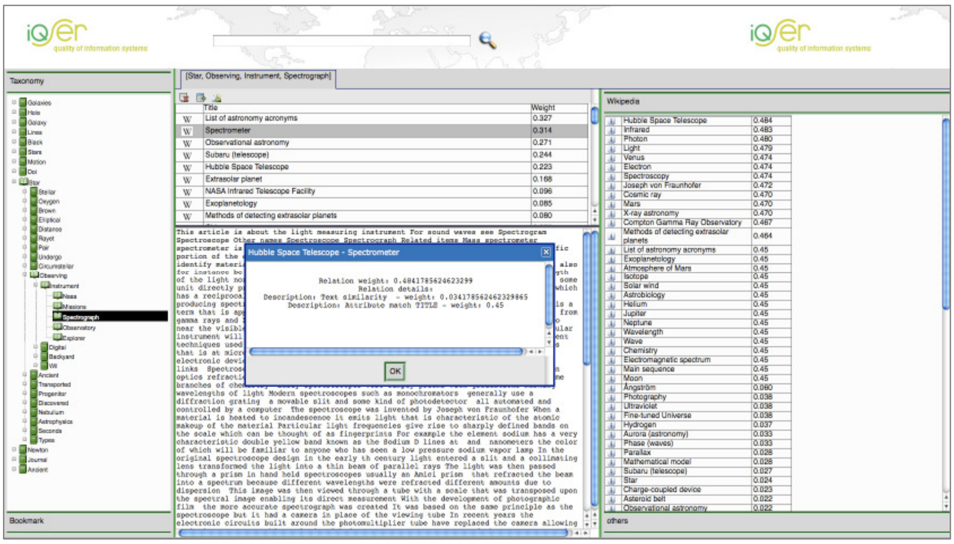

Fig. 2. The Wikipedia Example.

The shown application is a web client using ajax technology. The iQser middleware is used for integrating and analyzing the wikipedia articles. A user starts with an overview over the *main topics* inside a collection of articles by a concept tree with four levels (left side of the application). Each main topic is connected to *related concepts*. Users can now navigate to the fourth level on which they can see that "star" is one of the main concepts in the topic range "astronomy". Furthermore, users can recognize that one aspect of the concept "star" is "observation". In the context of "star" and "observation" one concept is "instruments". There are instruments to observe stars. In the context of the concepts "star", "observation" and "instruments", the tree shows "spectrograph" as example of an instrument to observe stars. By selecting one path in the tree the visitor of the demo can see clustered and ranked articles which are related to the concept constellation. The top articles are plausible allocations like "Spectrometer" and "Observational Astronomy". The algorithm for clustering is combined with an algorithm for a ranking that helps the user of the application to get a well organized overview. If the user of the application selects one article in the calculated cluster, then the extracted text from the html document of wikipedia appears and on the right site a list of connected wiki articles. If there would be more than one object type, they would be shown in separated accordion area for a better orientation. The related articles are ranked because of the calculated weight as a result from the combination of the weight of each found relations between two objects that is stored in an quadruple. A button at the left side of each article in the list opens a popup window with information about the reasons for the connection. They can be displayed in different ways. The shows just a summary but it is possible to show more detailed information like the matched concepts in a similarity calculation. Because of the non hierarchical structure the user of the application can use the connected articles to navigate through the article collection. If he would select one linked article, this article would appear in the center of the screen and at the right site its connections including the link back to the article he selected before.

Besides, the application also offers an explicit search field which can be used to perform a classical keyword-based search. If a user, for example, searches for "jupiter", he will be provided with a list of all articles about "Jupiter". Now, he can use this article as a starting point for further research without any need to set additional queries. An overview over related topics and opportunities to get deepening information is automatically provided. A user gets all relevant information in the given context. In the demo the context is an wikipedia article. In use cases for business a context could be any other business object characterizing the current focus of interest.

For another use case iQser developed a demo based on experiences of a project for a large system integrator. For this demo a CRM system, a file server and pages of a web domain are dynamically integrated in the platform to provide a central access of the distributed data. The semantic analysis connects all business objects regard to their content to support an account manager in his daily work. The test data was created for a fictive engeneering company for the automotive industry.

The screenshot shows on the left side the main navigation of the application. The dynamic calculated tree shows all types of business objects in the integrated data sources grouped by CRM data, files and webpages. The user can select one of this object types or start a full text search all over the integrated data by a search field on the top of the application.

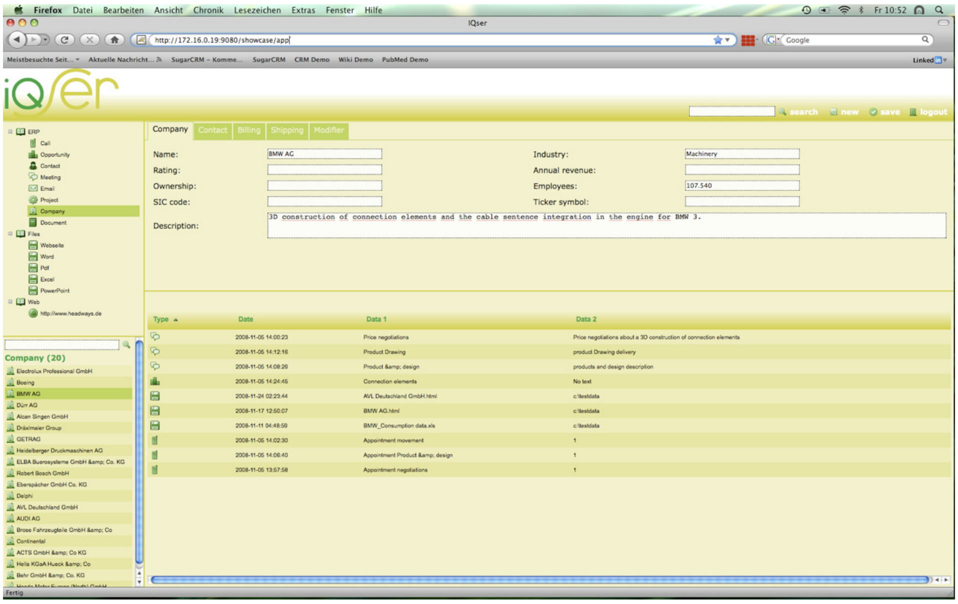

Fig. 3. The CRM Example.

The screenshot shows a result set of a selection in the navigation tree. Further search inside the result set are possible by a textfield on the left site under the navigation tree. Afterwards the user can make a selection in the result set like „BMW“ as it is shon in the screenshot to get all the detailed information about this company on the right site.

Under this company description of the integrated CRM system he see a list of connected business objects which are associated by the semantic analysis. This gives an account manager a 360° view over a customer to prepare the meetings, calls and campaigns. This overview includes all calls and meeting, which could come from different systems.

In thie given example the user can see related documents with background information about BMW but also about companies which are related to BMW by a supply chain. He can read the information by a single mouseclick right inside the application. The application also suggests one sales opportunity that has not be defined by anyone in a CRM system, bat that is found by analysing all the companies and opportuinities of the engeneering company.

Also other created connections between the business objects are usefull for any user of the system. The systems groups poeple with a similar profile to address them in a campaign for example. It also finds similar project for synergies in the development of a product or service. The best thing is, that the systems is updating itself near realtime, if any business objects are changing.

Because of the bidirectional integration of business objects the application can be used for procutive work. In the shown demo application iQser implemented afeature to modify or add new contacts by selecting the according buttons. Each new or modified information object is immediately updating the databse of the CRM system.

4 Discussion

The iQser middleware is available in a stable release. The architecture and algorithms have been proven in projects with T-Systems and EADS. By June 2009 a free *Software Development Kit* (SDK) will be available to create enterprise-specific solutions.

In particular, this SDK will enable enterprises to flexibly design and develop own applications based on the iQser middleware. Among other things, the SDK includes *Maven templates* to support efficient software development.

Besides, iQser offers a partner program including services like training of software developers and project managers. Also interested corporations can use the SDK to evaluate the platform for their requirements.

The benefits of the iQser approach can be summarized as follows:

- Saves time by automating information logistics that make search and manual organization of information unnecsessary.
- Enables the discovery of knowledge out of available data by analyzing unstructured information and by connecting available data and inherit concepts.
- Saves costs by a reducing complexity for application integration, application development and implementing services for process execution.

5 Related Work

For establishing the concept graph, iQser used research results of text mining, especially text segmentation and statistical methods, e.g. for a significance calculation of terms and co-occurences [4]. *Natural Language Processing* (NLP) also influenced the development of the iQser technology. In most cases, NLP approaches require a lot of work in training phases [5] or exhibit a growing complexity for the data model and cpu ressouces [6]. Inspiration for iQser have been use cases for an enhanced business intelligence, for example market research by analysing customers opinion [7]. The vision of Semantic Web is covered by the iQser approach to connect content across different applications and sources as well as options for the user to annotate, modify, recommand, commend the information or start any transaction. In opposite to the recommendations of W3C the iQser GIN Platform is not limited on standards for informations objects like RDF and also does not need any manual classification or ontology written in RDFS or OWL [8, 9]. But users can combine this approaches by the APIs of the platform for any data source or additional anayisis. Once a network or graph of information objects is established, the knowledge can be querried to answer questions and infer new knoweldge our of the given and analyzed data sources [10]. A query for the graph can be formalized as a graph itself to answer complex questions or generate reports [11]. Use cases for the Semantic Web, especially for konwledge management, knowledge visualization and knowledge discovery Are described in [12, 13]

6 Summary & Future Work

This paper has presented the iQser approach, a new semantic technology based on a powerful middleware platform (called iQser GIN platform). This practically-approved middleware platform enables enterprises to efficiently develop effective semantic applications integrating various sources of structured and unstructured information. Goal is to empower enterprise users to gain new insights into complex information domains. In order to illustrate the practical relevance of the approach, we have also discussed potential enterprise computing use cases such as business (process) intelligence, the alignment of process-aware information systems, and information retrieval.

Future work will deal with the development of further prototype applications using the iQser software development kit. Thereby, focus will be the support of complex and knowledge-intense enterprise domains such as the automotive industry.

References

1. Bela Mutschler, Manfred Reichert, Johannes Bumiller: "*An Approach to quantify the Costs of Business Process Intelligence*". Int'l. Workshop on Enterprise Modelling and Information Systems Architectures (EMISA '05), Lecture Notes in Informatics, Vol. P-75, pp. 152-165, Klagenfurt, Österreich, 2005.

2. D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, and M. Shan. *Business Process Intelligence*. Computers in Industry, 53, pp. 321-343, 2004.

3. Y. L. Antonucci. *Using Workflow Technologies to improve Organizational Competitiveness*. Int'l. J. of Management, 14(1), pp.117-126, 1997.

4. W. M. P. van der Aalst. *Business process management—A Personal View*. BPM J., Vol. 10, No. 2, pp. 248–253, 2004.

5. Ana-Maria Poperseu, Oren Etzioni. *Extracting Product Features and Opinions from Reviews*. in: Anne Kao, Stephen R. Poteet (Eds). Natural Language Processing and Text Mining, Springer 2007, pages 9 - 28.

6. Mauritius A. R. Schmidtler, Jan W. Antrup. *Automatic Document Separation: A Combination of Probabilistic Classification and Finite-State Sequence Modeling*. in: 2, pages 123 - 144

7. Gerhard Heyer, Uwe Quasthoff, Thomas Wittig. *Text Mining: Wissensrohstoff Text. Konzepte, Algorothmen, Ergebnisse*. W3L-Verlag 2006

8. Giovanni Marchisio, Navdeep Dhillon et al. *A Case Study in Natural Language Based Web Search*, in: 2, pages 69 - 90

9. Parcal Hitzler, Markus Krötsch, Sebastian Rudolph, York Sure. *Semantic Web. Grundlagen*. Springer 2008

10. Gabriele Wichmann. *Entwurf Semantic Web: Entwicklungen, Werkzeuge, Sprachen*. VDM Verlag Dr. Müller 2007

11. Grigoris Antoniou, Uwe Aßmann, Christina Barglio, Stefan Decker, Nicola Nenze, Paula-Lavinia Patranjan, Robert Tolksdorf (Eds). *Reasoning Web*. Third International Summerschool 2007, Springer 2007

12. Jörg Wurzer, *New Approach for Semantic Web by Automatic Semantics*. European Semantic Technology Conference, Vienna 2008.

13. Tassilo Pellegrini, Andreas Blumauer (Eds). *Semantic Web: Wege zurvernetzten Wissensgesellschaft*, Springer 2006

14. Gerold Riempp. *Integrierte Wissensmanagementsysteme. Architektur und praktische Anwendung*. Springer 2004