ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2007, Vol.36, No.1A

SCENARIO-DRIVEN INFORMATION RETRIEVAL: SUPPORTING RULE-BASED MONITORING OF SUBSEA OPERATIONS

Darijus Strasunskas, Stein L. Tomassen

Dept. of Computer and Information Science, Norwegian University of Science and Technology
NO-7491 Trondheim, Norway

Abstract. The production systems used by the subsea petroleum industry are knowledge and information intensive. Any problem needs to be solved quickly and efficiently avoiding decommissioning or waiting for the symptoms to be escalated. This requires precise information to be supplied on-time. For this reason we have proposed rule-based monitoring of device performance. However, covering all possible cases by rules is a labour-intensive and not trivial task. Therefore, in this paper we propose a scenario-driven information retrieval approach to complement rule-based monitoring. The main objective is to automatically formulate a query that is sent to a vector-space model information retrieval engine every time incomplete inference happens, i.e. when a specific case has no rules defined.

Keywords: Semantic technology, ontology, rules-based inference, information retrieval, integrated operations.

1. Introduction

An industry-driven consortium launched the Integrated Information Platform project [4, 11] in 2004. The project's primary objective is to extend and formalize an existing terminology standard for the petroleum industry, ISO 15926 [7]. Using OWL Full sublanguage, this standard is transformed into a real ontology that provides a consistent unambiguous terminology for subsea petroleum production systems. The ontology is used in monitoring of drilling and production processes.

The production systems used by the subsea petroleum industry are knowledge and information intensive. When a well is put into operation, the production has to be monitored closely to detect any deviation or problems. Any problem needs to be solved quickly and efficiently avoiding decommissioning or waiting for the symptoms to be escalated. Operators' task is actually even more complicated since analysis of a particular problem may involve hundreds of potential causes and require the consultation of a large number of documents.

Therefore, in this paper we propose a scenario-driven information retrieval approach that complements rule-based condition monitoring of subsea devices. The objective of this paper is to elaborate on task-specific information retrieval and how it can be integrated to rule-based system in order to support incomplete inference, employing scalability and efficiency of vector space retrieval engines.

The paper is structured as follows. Next we introduce the IIP project. Later we describe a motivating scenario for our approach. Then we elucidate our approach to integration of rule-based notification and task-specific information retrieval. Before concluding the paper, we overview related work.

2. The IIP Project

The Integrated Information Platform (IIP) project is a collaboration project between companies active on Norwegian Continental Shelf and academic institutions, supported by the Norwegian Research Council. Its long-term target is to increase petroleum production from subsea systems by making high quality real-time information for decision support accessible to onshore operation centres.

The IIP project [4] addresses the need for a common understanding of terms and structures in the subsea petroleum industry. The objective is to ease the integration of data and processes across phases and disciplines by providing a comprehensive unambiguous and well accepted terminology standard that lends itself to machine-processable interpretation and reasoning. This should reduce risks and costs in petroleum projects and indirectly lead to faster, better and cheaper decisions.

3. Illustrative Scenario

Consider a production operator monitoring the production efficiency of a well in the area of oil and gas exploration and production. She is located in a control room with several monitors showing the status of the wells. In such a control room, there are constant alarms of some sort with varying degree of importance. One of the most important responsibilities of the production operator is to look for tendencies among these alarms. One or more of these alarms can indicate an upcoming serious problem that might be handled in advance and hence avoiding a potential disaster. If she can lower the risk of these potential problems by acting quickly to those relevant alarms, the production can continue smoothly. Therefore, retrieval of the right information at the right time is an essential task here.

Continuing the scenario, consider the production engineer noticing a tendency of alarms pointing that temperature at choke inlet is increasing. Therefore, she has to find out diagnosis and a solution to this problem. On one of her many displays she sees that one of the alarms is related to the choke that is a part of a "*christmas tree*" installation, i.e. a component found among subsea equipment (see Figure 2c, visualization of the concepts/equipment classes related to "*christmas tree*"). She searches for possible cause and dependent measures in order to find a diagnosis and feasible solution to the problem.

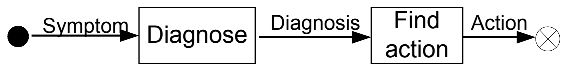

Figure 1. Illustrative activity

A simplified scenario for exemplification of the illustrational case and our approach is denoted in Figure 1. There "*Diagnose*" and "*Find action*" are the main tasks. An actual activity is more complicated [6] involving data pre-processing, mapping to ontology classes, tendency analysis, etc. However, for exemplification purpose we adopt simplified process containing most troublesome tasks for full automation.

4. An approach to scenario-driven information monitoring

There are envisioned several application areas of the subsea oil and gas production ontology. Interoperability in the highly multidisciplinary petroleum industry is the main goal [4], while the tasks of ontology-driven information retrieval [15] and rule-based notification [14] have main focus when it comes to supporting routine operations by information retrieval (IR). The rule-based approach is mainly applied to condition monitoring of subsea production. However, not all possible cases can be encoded in rules before hand. Furthermore, here information retrieval should be adjusted to the scenario, since precision of the retrieved information is very important. Therefore, here we present an approach to complement rule-based monitoring with a task-specific and ontology-based information retrieval.

Next we shortly introduce rule-based reasoning for condition monitoring followed by more detail discussion on a task-specific information retrieval. We elaborate on main principles and components of the integrated system.

4.1. Rule-based monitoring

A full case of condition monitoring consists of three main steps [14]: *Data processing*, *Health assessment* and *Treatment planning*. The data processing step takes care of analysis of data streams (Figure 2a, illustrates Daily Product Report - DPR) and mapping the actual measurements to data model (the ontology based on ISO 15926 and other standards regulating the petroleum domain). The output of this step is a detected state of equipment, for instance, an increased temperature measured at choke inlet, i.e. identification of symptom.

Having an identified tendency (symptom), next step is *health assessment*, i.e. inference of diagnosis. This step is heavily based on the rules and involves most of reasoning. The rules are used to identify possible causes, infer a diagnosis and finally lead to an action (treatment). At this step we employ rules defined in SWRL (Semantic Web Rule Language) [5]. For instance, *if a choke has a temperature sensor and temperature is equal or above the maximum operating temperature then the choke is in critical state*. This rule is illustrated below using SWRL built-in predicate `swrlb:greaterThanOrEqual` [5], and incoming data in XML format are exemplified in Figure 2a, measure class definition in Figure 2b. Then the rule defining dependencies among measurement classes is used to infer diagnosis, as follows.

hasTempera tureSensor (?x,?y) ∧ *hasTemp(?y ,?temp)* ∧
hasMaximum OperatingT emp(?x,?ma xtemp) ∧
swrlb:grea terThanOrE qual(?temp ,?maxtemp)
→ *inCritical State(?x,? temp)*

The *treatment planning* step takes care of the last two activities in the condition monitoring cycle, i.e., maintenance planning and actions that need to be taken in order to resolve the situation. This step either notifies the responsible controller who needs to perform the actions (e.g. *increase choke opening by 10%*) or executes the action automatically.

4.2. Scenario-driven information retrieval

In order to complement rule-based monitoring, we propose a scenario-driven information retrieval that is evoked every time incomplete inference happens, i.e. when a specific case has no rules defined. The main objective is to automatically formulate a query that is sent to a vector-space model information retrieval engine. Consequently query should be adjusted to corresponding tasks. In this subsection, we will first describe the information and knowledge resources that enable us to formulate task-specific queries and then we will present the scenario-driven information retrieval procedure.

For this purpose we adapt our method for the ontology-driven information retrieval [15] to support rule-based process of production monitoring. The idea [15] is to construct a *feature vector (FV)* for each of the concept defined in ontology. The feature vector is used to align a concept to a terminology of documents collection and later is used for query refinement. This is done by exploiting the ontological structures (i.e. semantic relationships a particular concept is involved in ontology) and computing statistical co-occurrence of words that are related to concepts in documents collection. The basic idea of FV is driven by use of a scalable vector-space retrieval engine.

Concepts and relations between them that are specified in ontology are used as starting point in indexing the document collection and extracting most relevant terms. These relevant terms constitute the basis for the concept feature vector. These terms are closely co-located and often appear together with a particular concept from ontology. The process of FV construction is elaborated in [16]. However, here we exemplify how a task-specific feature vector is created.

As said, feature vectors provide interpretations of the concepts with respect to the document collection. Synonyms and conjugations would naturally go into such a vector, but also related terms that tend to be used in connection with the concept are included to provide a contextual definition of it. This allows us to relate the concepts defined in the ontology to the terms actually used in the document collection.

Having the ISO 15926 standard specified as ontology, we relate discipline- and task-specific terminology to domain concepts. Each task has a term denoting its scope and, partially, a goal. For instance, the task "*Diagnose*" (Figure 1) has a goal to find a cause and diagnosis for a particular symptom. Therefore, we take this task-specific term (concept), and expand it by adding related terms from the thesaurus for the oil industry. In this case, adding terms and phrases as "reason, problem source, origin of problem, cause, etc." This set of related terms is used as main input for computing a task-specific feature vector.

Figure 3 illustrates the main components used in construction of the task-specific feature vectors, while more detailed FV construction process is described in [16]. Here, scenarios and related task-specific terms are extracted from a workflow repository, and expanded by a set of related terms (mainly using synonyms, hypernyms and hyponyms) from oil industry thesaurus. Then, task-specific feature vectors (FV_t) are computed for each pair <c, t>, where *c* is a concept name (e.g., from the IIP ontology, see Figure 2c for exemplification of the ontology) and *t* is a task-specific term. Task-specific feature vectors are built based on statistical co-occurrence of a task-specific terminology together with the concepts from ontology.

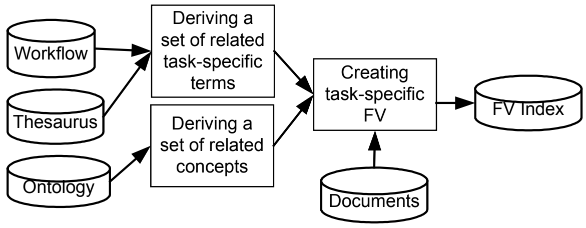

Figure 3. Main components in task-specific feature vector computation

Consider an experience report as follows[1], where underlined are statistically significant co-occurrence of terms related to "*choke*", while bold font emphasises the terms related to the tasks (e.g., action – emergency shutdown, halt; diagnosis – form): "*In the summer of 2004, the gas flowline was operated with the subsea choke wide open, controlling the flowline with the topside choke (to control slugging). The manifold pressure was nearly 4,100 psia. In mid-July, an* ***emergency shutdown (ESD****) was tripped, shutting in the Mica flowlines at the topside boarding valves. […] methanol injection at the manifold was started and the boarding choke was opened to blowdown the flowline (as per normal startup procedure). Approximately 2 hours after the blowdown was initiated, the subsea choke was opened to start production from the gas well, and as a result, the manifold pressure almost immediately increased 800 psi. In retrospect, this may have been an indication that a hydrate plug had* ***formed*** *and that all operations should have been* ***halted*** *for further engineering review.*"

Then possible task-specific feature vector for a pair <manifold pressure, action>[2] is as follows: {choke, manifold pressure, blowdown, emergency shutdown, ESD, halt, methanol injection>.


```xml
<witsml:facility>
  <witsml:name kind="wellhead" namingSystem="EnergyComponents">..
  </witsml:name>
  <witsml:facilityParent1 kind="well" namingSystem="EnergyComponent..
  </witsml:facilityParent1>
  <witsml:facilityParent2 kind="template" namingSystem="EnergyCompo..
  </witsml:facilityParent2>
  <witsml:unit>ASG-A_L-3H_wellhead</witsml:unit>
  <witsml:contextFacility kind="well" namingSystem="EnergyComponent..
  </witsml:contextFacility>
  <witsml:flow>
    <witsml:name>ASG-A_L-3H_wellhead_production</witsml:name>
    <witsml:kind>production</witsml:kind>
    <witsml:port>L-3H_wellhead_outlet</witsml:port>
    <witsml:qualifier>allocated</witsml:qualifier>
    <witsml:temp uom="degC">116.95241</witsml:temp>
    <witsml:pres uom="bar">147.76852</witsml:pres>
    <witsml:portDiff>
      <witsml:port>ASG-A_L-3H_portdiff</witsml:port>
      <witsml:presDiff uom="bar">45.54977</witsml:presDiff>
      <witsml:tempDiff uom="degC">5.83645</witsml:tempDiff>
      <witsml:chokeRelative uom="%">67.48616</witsml:chokeRelative>
    </witsml:portDiff>
  </witsml:flow>
</witsml:facility>
```


Figure 2a. A fragment of Daily Production Report in XML[3]

[1] Retrieved from Society of Petroleum Engineers, *http://www.spe.org/*.

[2] Here for simplification purposes, term weight is assumed to be equal.

[3] Here, WITSML – Wellsite Information Transfer Standard Markup Language, see *http://www.witsml.org/*.

```
<Class ID="ABD134">
  <subClassOfresource="&iso15926-4;Choke"/>
    <iso15926-4:maximumOperatingTemperature>
      <iso31:Temperature>
        <iso1000:celsius>
              300.0
        </iso1000:celsius>
      </iso31:Temperature>
    </iso15926-4:maximumOperatingTemperature>
etc.
</Class>
```

Figure 2b. Definition of maximum operating temperature for choke

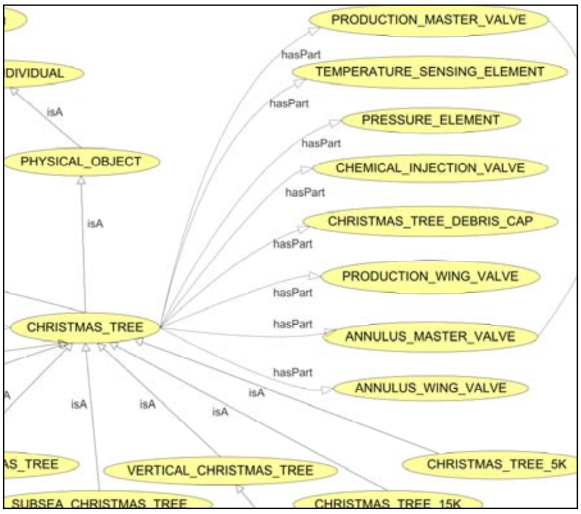

Figure 2c. A fragment of subsea oil and gas production ontology, based on ISO 15926

However, there is a great challenge to completely define rules for all possible dependencies between measures and corresponding actions needed to take to resolve the problematic situations. Operation controller can always refer to manuals or search in a document repository. However, switching between systems or changing the working way requires a considerable amount of time. Therefore, it is a desirable extension of the current systems to tightly integrate rule-based condition monitoring with information retrieval.

4.3. Interplay between rules and information retrieval

Interaction of the rule-based condition monitoring and notification with ontology-driven information retrieval system is shown in Figure 4. Here searching for relevant information is designed to be supplemental way of interaction with the rule-based system, since covering all possible cases by rules is a labour-intensive and not trivial task. Therefore, it is important to enable users to access previous reports and documents related to the problem on-hands. Smooth transition between these two different interaction ways is a challenge as well. Therefore, we propose an automatic query formulation based on a corresponding inference task that cannot be executed or a returned answer is incomplete.

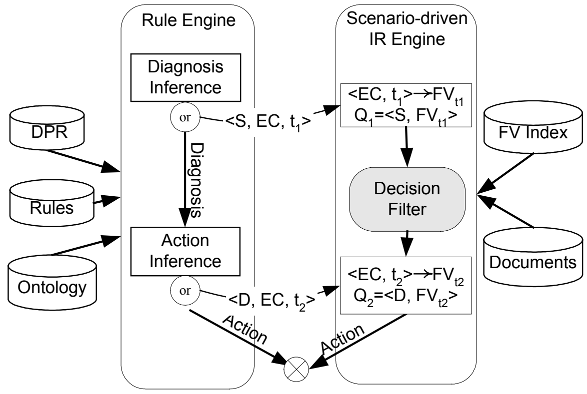

Figure 4. Procedure of interplay between rule and IR engine

A rule engine receives data from Daily Production Reports (DPR), uses rules and ontology to reason about a situation on-hands. If a rule is incompletely defined and no answer can be inferred, then the rule engine sends a triple <*input*, *equipment classes*, *task*> to a scenario-driven information retrieval engine. Here an *input* is incoming data to be used in a particular inference task. In Figure 4, "*Diagnosis inference*" uses a set (*S*) of symptoms (e.g. increasing pressure), while "*Action inference*" receives diagnosis (if any) as an input. For instance, after unsuccessful inference of diagnosis, symptoms, related equipment classes (concepts from ontology) and task (task name) are sent to IR engine (see Figure 4).

Then scenario-driven IR engine refines provided triple and expands the query using corresponding task-specific feature vectors (i.e. FV_{t1} is selected based on the provided concepts (*EC*) and task (t_1)). A component, called *Decision Filter*, has a function to extract a decision from the manually selected document. Actually, the selected relevant document is processed in a similar way as it is done while constructing task-specific feature vectors. Just here it is done locally by taking into account only the selected document, i.e. local vs. global document analysis [18]. Here, the first task-specific feature vector (FV_{t1}) is filtered out and reduced to the terms found in the selected document, i.e. $D \subseteq FV_{t1}$.

Query (Q_2) in a second task (finding an action) is formed as Q_1. First part is a set of diagnosis related terms (D) received either from the rule engine (assuming termination of reasoning after successful diagnosis inference), or from the previous scenario-driven IR task. Second part is expansion of EC and t_2 by a task-specific feature vector (FV_{t2}) as in Q_1.

5. Related work

The problem described here could perhaps be solved using other technologies. For instance, applying fuzzy expert systems and fuzzy reasoning [12] or non-monotonic reasoning [1], that is suitable for reasoning in the cases of incomplete information and knowledge as well inconsistent information.

However, the Norwegian oil industry decided to rely on the Semantic Web technology as a platform for future integrated operations. This comes along with benefits such as semantic interoperability, common inter-disciplinary terminology, etc. Here we have focused on how to support the underlying information platform.

Liu & Chu [8] have proposed an approach to knowledge-based query expansion to support scenario-specific retrieval of medical documents. Their approach is most similar to ours as they use both statistical co-occurrence and domain knowledge in order to expand the query. However, they rely only on concepts co-occurrence; while we do take into account other terms collocated with a concept of interest. Furthermore, they derive scenario-specific concepts from a knowledge base, namely UMLS[4] (Unified Medical Language System). They use semantic network to identify scenario-specific concept relations, for instance, having specified that *a medical device and pharmacological substance treat disease*, they are able to identify the semantic type that a concept belongs to and in this way relate concepts "*contact lens*" and "*keratoconus*"[5] to a scenario that is "*treatment*". Contrary to them, our approach is based on explicitly defined activities (workflows), where we extract a task-specific terminology and construct task-specific feature vectors for each concept.

A different approach is chosen by members of the Aksio project [9]. They propose a process driven approach to access experience from daily drilling reports. However, they rely on experts' annotating the reports and use only ontology concepts and relations between them to expand query. Skalle & Aamodt [13] propose a combined reasoning method (using case-based and model-based reasoning) to support decision in fault diagnosis in oil well drilling.

Furthermore, an important body of work exists in query expansion area (e.g. [2, 10, 17, 18]). Most query enrichment approaches are not using ontologies like [2, 3, 10]. Query expansion is typically done by extending provided query terms with synonyms or hyponyms. Qiu and Frei [10] are using query expansion based on similarity thesaurus. Similarly, Grootjen and van der Weide [3] describe a conceptual query expansion. There, the query concepts are created from a result set. Chang, Ounis and Kim [2] is also not using ontologies but is reliant on query concepts. Two techniques are used to create the feature vectors of the query concepts, i.e. based on document set and result set of a user query.

6. Conclusions

The Integrated Information Platform project is one of the first attempts at applying state-of-the-art Semantic Web technologies in an industrial setting. With the ISO 15926 ontology at hand, the industry will have taken the first step towards integrated operations on the Norwegian Continental Shelf. Data can then be related across phases and disciplines, helping people collaborate and reducing costs and risks.

One of the applications developed in IIP is a system for ontology-driven task-specific reasoning and information retrieval. In this paper we presented an approach to task-specific information retrieval to complement rule-based notification. Here, the concepts in the ontology are associated with contextual task terminology in terms of feature vectors tailoring the ontology to the content of the document collection. This adaptation is fundamental in order to provide useful and usable services to a variety of users in the presence of large variations in resources and activities. Further, the feature vector is used to enrich a provided query. Query enrichment by task-specific feature vectors provides means to bridge the gap between query terms and terminology used in a document set, and still employing the knowledge encoded in ontology.

The main advantage of the proposed approach is integration of structured data and knowledge with unstructured information (documents in natural language). However, as future work, we will need to experimentally validate our approach in a bigger scale. Possible future extensions of the approach would include an experimenting with semantic web services and more tight integration of reasoning and information retrieval. In a current version of the approach there is only one-way communication between rule engine and IR engine. While reasoning on information retrieved from documents could bring additional advantages.

Acknowledgement

This research work is funded by the Integrated Information Platform for reservoir and subsea production systems (IIP) project, which is supported by the Norwegian Research Council (NFR), project number 163457/S30.

References

[1] **G. Brewka.** Nonmonotonic reasoning: Logical foundations of commonsense. *Cambridge University Press, Cambridge*, 1991.

[2] **Y. Chang, I. Ounis, M. Kim.** Query reformulation using automatically generated query concepts from a document space. *Information Processing and Management* 42, 2006, 453-468.

[3] **F.A. Grootjen, T.P. van der Weide.** Conceptual query expansion. *Data & Knowledge Engineering* 56, 2006, 174-193.

[4] **J.A. Gulla, S.L. Tomassen, D. Strasunskas**. Semantic Interoperability in the Norwegian Petroleum Industry. *In D. Karagiannis, H.C. Mayer (eds.), 5th International Conference on Information Systems Technology and its Applications* (ISTA 2006), *Vol.*P-84, *Kollen Druck+Verlag GmbH, Bonn*, 81-94.

[4] *http://umlsinfo.nlm.nih.gov/.*
[5] An eye disease.

[5] **I. Horrocks, P. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean.** SWRL: A Semantic Web Rule Language. Combining OWL and RuleML. 2004. *Retrieved: March 5, 2007 from http://www.w3.org/ Submission/SWRL/.*

[6] ISO 13374. Condition monitoring and diagnostics of machines. Data processing, communication and presentation. 2003. *Retrieved: March* 5, 2007 *from http://www.iso.org/iso/en/CatalogueDetailPage. CatalogueDetail?CSNUMBER=21832*.

[7] ISO 15926-2. Integration of life-cycle data for process plants including oil and gas production facilities – Part 2: Data model. 2003. *Retrieved: March* 5, 2007 *from http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail? CSNUMBER=29557*.

[8] **Z. Liu, W.W. Chu.** Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *Information Retrieval* 10(2), 2007, 173–202.

[9] **D. Norheim, R. Fjellheim.** AKSIO – Active Knowledge management in the petroleum industry. *Proceedings of the ESWC* 2006 *Industry Forum*, Montenegro, 2006.

[10] **Y. Qiu, H.-P. Frei.** Concept based query expansion. *Proceedings of the* 16*th international ACM SIGIR conference on Research and development in information retrieval. ACM Press, USA*, 1993, 160-169.

[11] **N. Sandsmark, S. Mehta.** Integrated Information Platform for Reservoir and Subsea Production Systems. *Proceedings of the* 13*th Product Data Technology Europe Symposium* (PDT 2004), *Stockholm*.

[12] **W. Siler, J.J. Buckley.** Fuzzy Expert Systems and Fuzzy Reasoning. *John Wiley & Sons, Inc*., 2005.

[13] **P. Skalle, A. Aamodt.** Knowledge-based decision support in oil well drilling. *In: Intelligent Information Processing* II. *IFIP International Conference on Intelligent Information Processing, Beijing, China: Springer Science+Business Media B.V.* 2004, 443-455.

[14] **D. Strasunskas.** Resource Monitoring and Rule-Based Notification. Applications in Subsea Production Systems. *In Proceedings of the* 2007 *IRMA International Conference on Managing Worldwide Operations and Communications with Information Technology* (IRMA 2007), *Vancouver, Canada*, 2007, *IDEA Group Publishing*.

[15] **S.L. Tomassen, J.A. Gulla, D. Strasunskas.**. Document Space Adapted Ontology: Application in Query Enrichment. *In Kop, C. et al. (eds.),* 11*th International Conference on Applications of Natural Language to Information Systems* (NLDB 2006), *LNCS* 3999*, Springer-Verlag*, 46-57.

[16] **S.L. Tomassen, D. Strasunskas.** Query Terms Abstraction Layers. *In Meersman, R. et al. (eds.): OTM Workshops* 2006, *LNCS* 4278. *Springer-Verlag*, 2006, 1786-1795.

[17] **E.M. Voorhees.** Query expansion using lexical-semantic relations. *Proceedings of the* 17*th international ACM SIGIR conference*, 1994, 61-69.

[18] **J. Xu, W.B. Croft.** Query expansion using local and global document analysis. *Proceedings of the* 19^{th} *International ACM SIGIR Conference*, 1996, 4-11.

Received March 2007.