Integrating different grain levels in a medical data warehouse federation

Marko Banek[1*], A Min Tjoa[2], and Nevena Stolba[3]

[1] FER, University of Zagreb, HR-10000 Zagreb, Croatia
marko.banek@fer.hr
[2] Institute of Software Technology and Interactive Systems,
Vienna University of Technology, Favoritenstr. 9-11/188, A-1040 Wien, Austria
amin@ifs.tuwien.ac.at
[3] Women's Postgraduate College for Internet Technologies
Vienna University of Technology, Favoritenstr. 9-11/188, A-1040 Wien, Austria
stolba@wit.tuwien.ac.at

Abstract. Healthcare organizations practicing evidence-based medicine strive to unite their data resources in order to achieve a wider knowledge base for sophisticated research and matured decision support service. The central point of such an integrated system is a data warehouse, to which all participants have access. In order to insure a better protection of highly sensitive healthcare data, the warehouse is not created physically but as a federated system. The paper describes the conceptual design of a health insurance data warehouse federation (HEWAF) aimed at supporting evidence-based medicine. We address a major domain-specific conceptual design issue: the integration of low-grained, time-segmented data into the traditional warehouse, whose basic grain level is higher than that of the time-segmented data. The conceptual model is based on a widely accepted international healthcare standard. We use ontologies of the data warehouse domain, as well as of the healthcare and pharmacy domains, to provide schema matching between the federation and the component warehouses.

1 Introduction

Evidence based medicine [14] successfully weaves the clinical decision process based on human knowledge and the most efficient and most accurate computer-supported research evidence. The application of its concepts speeds up the transfer of clinical research findings into practice, which finally leads to cost reduction both for patients and health insurance organisations, as well as the improvement of the healthcare process as whole. A data warehouse storing a huge amount of healthcare data is the central part of an information system that supports evidence-based medicine.

---

* The scholarship at the Institute of Software Technology and Interactive Systems, Vienna University of Technology was funded by Austrian Federal Ministry for Education, Science and Culture under the grant "Ernst Mach"

Healthcare institutions nowadays join their data into a single warehouse in order to achieve a broader and more comprehensive data foundation for knowledge discovery, to the benefit of all participants. However, laws insuring privacy protection forbid the confidential healthcare data to be copied and distributed outside the healthcare organizations. Instead of copying data physically into a new warehouse, a logical integration, a federated data warehouse [15] is created.

In this paper we propose a multidimensional conceptual model of a federated data warehouse for the purpose of evidence-based medicine. The conceptual model of the component warehouses offers a traditional view on financial measures, yet it does not enable the processing of time-segmented medicine administration data (an important topic in evidence based-medicine), whose grain level is even lower than the basic grain level of the model. The contributions of our paper are the following: (1) we develop a federated conceptual model that successfully integrates the low-level, time segmented data but keeps the higher basic grain level. (2) Since the medicine administration quantities can be summed (like measures) and used as aggregation criteria (like dimensions) they behave as a "cube in cube". We regard the time-segmented data as a unique XML structure, extending the existing approaches for merging OLAP systems with XML documents described in [12, 13].

The paper is structured as follows. In Section 2 the requirements to a warehouse federation in healthcare domain are explained. Section 3 presents how the integration of different grain levels into the federated model is achieved. The use of ontologies in matching schemas for the federation is described in Section 4. An outline of the related work is given is Section 5. Finally, in Section 6 conclusions are drawn.

2 Federated Data Warehouse for Healthcare

A federated database is "a collection of cooperating database systems that are autonomous and possibly heterogeneous" [15]. The existence of a federation must not have any impact on the local users of a component database. A federated data warehouse is a functional data warehouse, a "big umbrella" [4]. No central, large data warehouse collecting data from smaller component warehouses is created: heterogeneous data warehouses are integrated into one unit from the conceptual point of view. A "common business model" [4] (i.e. common conceptual model) is needed, which defines common facts and dimensions.

The paper describes how data warehouses of different health insurance organisations in Austria are merged in an evidence-based medicine collaboration project. The case study of the federated data warehouse is called HEWAF (<u>He</u>althcare <u>Wa</u>rehouse <u>Fe</u>deration). A universal, simple and flexible common conceptual model is needed to enable potential future integrations to be done seamlessly and with a minimum effort. The conceptual model of HEWAF is based on the international healthcare standards HL7 [21] in order to achieve a high level of generalisation and portability. Version 3 of HL7 standards defines the object-oriented Reference Information Model (RIM, the starting point for all HL7 standards) and the Clinical Document Architecture (CDA), an XML-based document mark-up standard, which specifies the structure and semantics of clinical documents used for their exchange. Comparing to other existing standards, we claim HL7 to be the most comprehensive. Other standards either have a too strict format or allow too much semantic or structural ambiguity.

HEWAF will be used for OLAP queries and data mining. Therefore the conceptual model of the federation will be subject-oriented and multidimensional. Dimensions and facts should conform to HL7 RIM classes. Dimension attributes and measures should conform to attributes of HL7 classes and it must be possible to display each record (either at the basic-grain level of granularity, or an aggregated structure) as a CDA-conforming XML structure.

On the one hand, the federated warehouse will be used to find correlations between certain symptoms and diagnostic findings, to propose the best suitable therapy given the complete anamnesis status of a patient or to compare the efficiency of different medication products against the same disease. On the other hand, since health insurance organisations pay for therapy procedures and prescribed medications, they want to find possible unnecessary costs (e.g. medication therapies applied for a long time with no or very little effect) and to foster therapies that give the best possible efficiency for equal costs. Different from the patient-centred model based on the Electronic Patient Record, which is proposed in [9], our warehouse model is focused on three billable acts: patient encounters, therapies and prescriptions, which correspond to the HL7 RIM Act classes *PatientEncounter*, *Procedure* and *SubstanceAdministration*, respectively. In this way, all data relevant for a patient as well as financial data is captured.

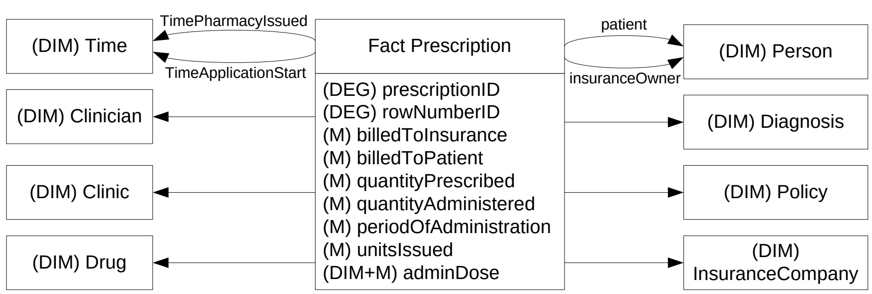

Fig. 1. The *Prescription* fact in HEWAF and its dimensions

The remainder of the paper describes the fact *Prescription* in the data warehouse star schema. The fact is characterized by eight standard dimensions (Fig. 1). The structures of the dimensions *Patient* and *InsuranceOwner* are identical, as they both provide a different view of personal data (if the treatment of a patient is carried by her/his own insurance policy, contents of both dimensions are the same). Analogously, there are two time dimensions of identical structure: *TimePharmacyIssued* and *TimeApplicationStart*. The value of the latter is often not exactly specified and in that case it is generally presumed to be equal to the value of the first. The other six dimensions are *Clinic*, *Clinician*, *Diagnosis* (only one, “main” diagnosis is specified in prescription documents), *Drug*, *Policy* and *InsuranceCompany*. There are two degenerate dimensions (i.e. single-attribute dimensions implemented as non-additive measures, [5]) corresponding to the serial numbers in the prescription document:

prescriptionID and *rowNumberID* (Fig. 1). The fact contains two financial measures, *billedToInsurance* and *billedToPatient*, and four additional measures: *quantityPrescribed*, *quantityAdministered* (the prescribed quantity can be 10 pills, but we can buy only a package of 12 pills), *periodOfAdministration* and *unitsIssued*. Finally, there is an additional fact attribute, *adminDose*, which contains a sequence of time-labelled segments giving a detailed description of medical substance administration.

3 Low-grained Medicine Administration Data

The fact *Prescription* gives us the ability to determine how different variants of prescription processes of the same drug substance can influence the patient's recovery and lead to a cost reduction. Several examples of the dosage instructions (also called *medicine administration* data) in the prescription document can be seen in Fig. 2.

1. Take 1 tablet each 8 hours for next 10 days
2. Take 600 mg each 12 hours for next 7 days
3. Take 2 tablets on the first day, and 1 on next four days (second tablet on the first day 12 hours after the first; tablet on the second day 12 hours after the last on the first day; later 24-hour periods)
4. Take 500 mg in the morning and 500 mg in the evening for next 7 days.

Fig. 2. Variants of dosage instructions (i.e. administration notes) in prescription documents

The total prescribed quantity (the quantity measure of the fact *Prescription*) is either expressed in units of weight or volume (for solid substances and liquids, respectively). Although it is possible to calculate the weight of the medical substance in a dilute solution, this is neither easy (we should collect data about concentrations of medical substance for all liquid drugs) nor practical (quantity is not additive across all levels of dimension *Drug*, similar to the number of units of sold goods in a grocery store [5]). Therefore the single attribute quantity consists of two parts: value and unit (which is in accordance with HL7).

All component data warehouses store administration instructions of prescriptions as a single attribute, *adminDose*, (Table 1, similar to the notation in Fig. 2), making them unsuitable for processing. Such a level of granularity also corresponds to the basic granularity of the fact *Prescription* in the federation model. However, the federation must enable administration data processing and putting queries at a detailed level. The grain level needed for their processing is lower than the basic grain level (i.e. the physical storage level) in the component warehouses. We argue that lowering the basic grain [5, 10] to the level of administration data is not an advisable solution for the following two reasons: (1) the semantics of the time dimension is different in the two cases, and fragmenting financial measures is meaningless; (2) the general focus of interest in the clinical decision making process is often mainly aimed at prescription data (as in Fig. 1), not at the instructions for medicine administration.

We propose a federated conceptual model whose basic grain corresponds to that of the component warehouses and to the basic federate schema in Section 2, but enables the integration of data at an even lower grain level. We can observe the medicine administration data as a complex attribute of XML type in an object-relational database (Fig. 3). It is a union of separate dosage records, time-labelled segments similar to a kind of a patient's diary. Each record (<*dose*>) refers to a separate dose, an act of drug consumption at a different point in time. Every separate dose, a <*dose*> sub-element of <*adminDose*>, contains a sequence number attribute, *sn*, two attributes defining the timestamp: *ts* (value) and *tu* (unit), and two attributes defining the quantity *q* (value) and *u* (unit).

This kind of representation entirely follows the idea adopted by HL7, with only one *doseQuantity* attribute, defined as a sequence of partial dosage records [21] . On the contrary, the CDA representation of *doseQuantity* does not view a dose as an inseparable, atomic unit, but splits time and quantity data, thus not being able to reach as much expressive power as our model.

Table 1. Part of the prescription fact table in a relational component warehouse

drug ID	billed-ToIns ($)	qtyPrsc	qtyAdm	prdAdm	noUnits	adminDose
1234	23,50	7500 [mg]	8000 [mg]	10 [day]	2	1 t each 8 hours
1234	23,50	7000 [mg]	8000 [mg]	7 [day]	2	2 t each 10 hours

```
<adminDose>
   <dose sn="1" ts="0" tu="h" q="250" u="mg"/>
   <dose sn="2" ts="12" tu="h" q="250" u="mg"/>
   <dose sn="3" ts="1" tu="d" q="250" u="mg"/>
   <dose sn="4" ts="2" tu="d" q="250" u="mg"/>
   <dose sn="5" ts="3" tu="d" q="250" u="mg"/>
   <dose sn="6" ts="4" tu="d" q="250" u="mg"/>
</adminDose>
```

Fig. 3. Medicine administration data in an XML notation suitable for processing

3.1 Queries containing medicine administration data

Following the concept of HL7, we perceive *adminDose* as an XML structure that is a single attribute of the HEWAF prescription fact. In this Section we show how the dosage quantities can be aggregated as measures and used as selection or aggregation criteria like dimensions. Hence, we call the *adminDose* attribute a "cube in cube".

Querying the medicine administration data and associating it with other warehouse attributes follows the principles of XML-extended OLAP querying elaborated by D. Pedersen, Jensen and T.B. Pedersen [12, 13]. In their approach, a standard OLAP fact can have one or more XML structures as dimensions. The approach is based on an extended OLAP algebra and XPath [17] structures are added to SQL to fetch XML data (the language is called SQL_{XM}). They distinguish three different ways XML data can be included into OLAP queries. First, queries may be *decorated* with XML data. Second, external XML data may be used for *selection* (operation of restriction in relational algebra). Third, OLAP data may be grouped by the values of external XML data when aggregation is performed.

We adopt the basic principles of this approach, with one substantial difference: our "XML" data is part of the fact (a "cube in cube"), not dimensions. Therefore we must differentiate the *internal* aggregation, which processes data within the same fact record, and the *global* aggregation, which is described in [12, 13]. In HEWAF, *decoration* and *selection* queries can include internally aggregated data. Aggregation operators SUM, COUNT, MAX, MIN, AVG can be used, except for the constraint that quantities measured by weight and those measured by volume cannot be aggregated together. In the following queries we extend the principles of SQL_{XM}.

Decoration. Decoration means providing supplementary medicine administration data in the result of an OLAP query, without using it as selection (WHERE clause of an SQL query) or global aggregation (GROUP BY clause) criteria.

In Query 1 a decoration without internal aggregation is presented. For each patient who was prescribed 250 mg azithromycin capsules in May 2005, textual description of the cause diagnosis should be displayed, as well as the quantity (value and unit) of the first dosage portion and region where the patient lives.

Query 1: SELECT d.description, **fp.adminDose/dose[@sn="1"]/@q**, **adminDose/dose[@sn="1"]/@u**, p.region FROM FactPrescription fp, Diagnosis d, Time t, Drug dr, Patient p WHERE Time.month="05_2005" and Drug.drug_code= "(azithr_250mg_code)"

Selection. In Query 2 medicine administration data is part of the WHERE clause, i.e. determines selection. Also, it is aggregated internally, using the operator SUM. The result returns the (altogether) cost billed to insurance institutions for all prescribed antibiotics where the dosage on the first day (first 24 hours) is smaller than or equal to the fifth of the entire prescribed dose. The results are grouped by diagnosis (the second level of ICD-10 classification, [16]) and diagnosis ID, as well as diagnosis description are displayed.

Query 2: SELECT SUM(fp.billedToInsurance), d.level2ID, d.level2Desc FROM FactPrescription fp, Diagnosis d WHERE **SUM(fp.adminDose/dose[@ts<"24" and @tu="h"]/@q)**/(fp.quantityPrescribed)<=0.2 GROUP BY d.level2ID

Aggregation. Aggregation can be meaningfully performed over attributes with a discrete domain of values. Considering time-segmented dose records, serial number (*sn*) has a discrete domain, while timestamp (*ts*) and quantity (*q*) are typical features with continuous values. In general, continuous domains should be split into a finite set of groups and hierarchies created over the groups in order to perform aggregations.

We state one hour as the shortest periods of time during which aggregation can be performed. Larger periods are 3-hour, 6-hour, 12-hour, 24-hour, 2-day, 7-day, 14-day and 4-week period. Considering the dosage portions that are at the border of these intervals, we define a time interval from *a* to *b* as [*a*, *b*), meaning that the portion at the lower border (with timestamp *a*) is included in the interval while the upper border (with timestamp *b*) is not included.

Although the aggregation over a single (non-grouped) quantity value seems possible but meaningless, for most of the products sold as pills, capsules or tablets there is actually only a finite set of possible administration quantities. Moreover, medical standards for administering liquids also prescribe a finite number of possible quantities. For instance, we can aggregate dosage data if we constrain the value of attribute *basicProductForm* in dimension *Drug* only to "capsule" and "tablet" and restrict the query selection (WHERE clause) to the three lowest levels of the *Drug* dimension hierarchy (a packed factory product: Sumamed 250g, Zithromax 500g, a factory product: Sumamed, Zithromax or a substance: Azithromycin).

While in the approach described in [12, 13] the external XML dimension is used to define aggregation criteria, not changing the essence of the aggregation process, in our case the basic granularity of the *Prescription* fact is split to the level of an individual administration dose and then the aggregation is performed.

Query 3 selects only prescriptions of 250 mg Zithromax capsules, groups them according to the dosage quantities (quantity is the concatenation of quantity value and quantity unit) and counts the quantities, so that we can see which dosage quantity is preferred.

Query 3. SELECT **concat (fp.adminDose/dose/@q, fp.adminDose/dose/@u),** COUNT(**fp.adminDose/dose/@q**) FROM FactPrescription fp, Drug d WHERE d.drug_code= "(zithromax_250mg_code)" and d.basicProductForm="capsule" GROUP BY **concat (fp.adminDose/dose/@q, fp.adminDose/dose/@u)**

4 Use of Ontologies in Achieving Federation

There have been many efforts during the last two decades to automate the process of integrating component database schemas, and, more recently, component warehouse schemas into a single federated schema. Sheth and Larson state that a completely automatic schema integration is not possible because of the inability of the semantic models to capture a real-world state completely, as well as the existence of multiple views and interpretations of a real-world state [15].

Recent research into ontologies and development of large lexical databases like WordNet [20] enables applications to "understand" the semantic relations between items. Ontologies organize terms of a certain domain into classes (e.g. wine and orange juice are both subclasses of drink and they are mutually disjoint) and state their relationships (e.g. colour is a property of wine). Lexical databases provide a thesaurus of synonyms, antonyms and homonyms.

Using ontologies and lexical databases in a rigorous manner could lead to a database federation schema matching process that would be almost fully automated. The application would deliver enough semantic information for the matching process, and the warehouse designer would only have to check its consistency. We briefly present the idea of an ontology-based warehouse schema matching, which is currently being implemented in W3C Web Ontology Language (OWL, [18]).

The role of ontologies in our approach is twofold. First, since there is no standard conceptual model of data warehouse, from which an ontology could be derived, we define our own “structural ontology” over the basic data warehousing constructs: cubes, dimensions, hierarchies, hierarchy levels, measures, dimensional attributes etc (the left side of Fig. 4). Ontology instances, which provide the structural schema of the federation and the component warehouses, can be automatically created from the warehouse metadata (the right side of Fig. 4). Second, additional ontologies must be used in order to match the structural components and accomplish the schema matching process. The ontologies correspond to domains described by the warehouse facts and dimensions. In our case we primarily need a healthcare and pharmacy ontology (which characterizes diseases, drugs, patients), but also other ontologies, referring to time (segments and units) or personal data. The use of WordNet should give a substantial support in building the ontologies. Moreover, WordNet will be used in cases when the query against ontologies produces no successful matching.

```
<owl:Class rdf:ID="Attribute"/>

<owl:Class rdf:ID="DimAttribute">
  <rdfs:subClassOf
          rdf:resource="#Attribute"/>
</owl:Class>

<owl:Class rdf:ID="Level" />

<owl:ObjectProperty
rdf:ID="attributeOfLevel">
  <rdf:type
rdf:resource="&owl;FunctionalProperty"/>
<rdfs:domain
rdf:resource="#DimAttribute"/>
    <rdfs:range  rdf:resource="#Level"/>
</owl:ObjectProperty>
```

```
<Level rdf:ID="ICDLevel2"/>

<DimAttribute
rdf:ID="level2Desc">
  <attributeOfLevel
rdf:resource="ICDLevel2">
</DimAttribute>

<Dimension
rdf:ID="Diagnosis">
<hasLevel
rdf:resource="ICDLevel1">
<hasLevel
rdf:resource="ICDLevel2">
<hasHierarchy
rdf:resource="DiagHierDef">
</Dimension>
```

Fig. 4. Ontology of the data warehouse domain

5 Related work

One of the first scientific papers describing an accomplished clinical data warehouse development project is [3]. Differences between conventional, business-oriented data warehouses and clinical data warehouses are outlined in [9] and key research issues in clinical data warehousing identified. The warehouse outlined in [9] has the patient in the focus of its interest, while the approach in [10] is similar to ours, focusing on billable acts (encounters).

Some resemblance to our approach can be seen in [6], which illustrates an association rule mining technique over human sleep datasets. There is a macro-level dataset containing patient demographics, general and sleep-specific habits and sleep-related disorders, as well as a micro-level dataset consisting of detailed, time-series data obtained from sensors (data such as electro-encephalogram or electro-myogram).

A data warehouse federation is described in [1, 8]. XML Topic Maps (XTM) [19] is used to provide the integration framework. Various elements of the federation model and the component warehouses are described as topic map resources, providing an abstraction layer and thus seamlessly overcoming the semantic heterogeneity. However, the topics must be manually defined first.

In [11] an enterprise knowledge portal that integrates OLAP and information retrieval (IR) functionality is presented, with an RDF-based ontology as the core, which stores metadata for different resources. A similarity function for metadata search is defined, which treats two identical descriptions as a 100% match, while two completely unrelated descriptions are a 0% match. Construction of a federated spatial database, using OWL, is outlined in [7]. Spatial databases are defined by using a particular XML encoding for the transport and storage of geographic information. The ontology is derived from WordNet and the spatial data transfer standard. A self-medication information system, which proposes to patients information and services on mild clinical signs and associated treatments, is illustrated in [2]. Given the simplified patient's electronic health record as input, an ontology is used to infer the right treatment proposal out of the self-medication knowledge base.

6 Conclusion

This paper presents the conceptual model of HEWAF, a federated data warehouse for health insurance organizations practicing evidence-based medicine. The model is based on the widely adopted international standard HL7. The biggest challenge we addressed during the design process was the integration of time-segmented, low-grained medicine administration data into the standard conceptual multidimensional data warehouse model with financial measures. The medicine administration data is a complex attribute of the fact, which can be observed as a "cube in cube" and aggregated internally. The basic principles of OLAP-XML federation are adopted and extended to generate queries. In order to automate the schema matching process as much as possible, we developed a data warehouse domain ontology, and are currently working on developing a simplified domain ontology in the healthcare and pharmacy area.

References

[1] Bruckner, R.M., Tok, W.L., Mangisengi, O., Tjoa, A.M.: A Framework for a Multidimensional OLAP Model Using Topic Maps. In: Proc. Int. Conf. on Web Information Systems Engineering (WISE'01). IEEE Computer Society (2001) 109-118.

[2] Curé, O.: Ontology Interaction with a Patient Electronic Health Record. In: Proc. Symp. Computer-Based Medical Systems (CBMS'05). IEEE Computer Society (2005) 323-328

[3] Ewen, E.F., Medsker, C., Dusterhoft, L.E., Levan-Shultz, K., Smith, J.L., Gottschall, M.A.: Data Warehousing in an Integrated Health System: Building the Business Case. In: Proc.

Int. Workshop on Data Warehousing and OLAP (DOLAP'98). ACM Press, New York (1998) 47-53

[4] Jindal, R., Acharya, A.: Federated Data warehouse Architecture, Wipro Technologies white paper, (2004) http://hosteddocs.ittoolbox.com/Federated%20data%20Warehouse%20Architecture.pdf (Last access: April 10, 2006)

[5] Kimball, R., Ross, M.: The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling. 2nd edn. John Wiley & Sons, New York (2002)

[6] Laxminarayan, P., Ruiz, C., Alvarez, S.A., Moonis, M.: Mining Associations over Human Sleep Time Series. In: Proc. Symp. Computer-Based Medical Systems (CBMS'05). IEEE Computer Society (2005) 323-328

[7] Morocho, V., Saltor, F., Pérez-Vidal, L.: Schema Integration on Federated Spatial DB across Ontologies. In: Proc. Int. Workshop on Engineering Federated Information Systems (EFIS'03). IOS Press (2003) 63-72

[8] Nguyen, T.B., Tjoa, A.M., Mangisengi, O.: MetaCube XTM: A Multidimensional Metadata Approach for Semantic Web Warehousing Systems. In: Proc. Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK'03). Lecture Notes in Computer Science, Vol. 2737, Springer Verlag, Berlin Heidelberg New York (2003) 76-88

[9] Pedersen, T.B., Jensen, C.S.: Research Issues in Clinical Data Warehousing. In: Proc. Int. Conf. on Scientific and Statistical Database Management (SSDBM'98). IEEE Computer Society Press (2001) 43-52

[10] Song, I.-Y., Rowen, W., Medsker, C., Ewen, E.F.: An Analysis of Many-to-Many Relationships between the Fact and Dimension Tables in Dimensional Modeling. In: Proc. Int. Workshop on Design Management of Data Warehouses (DMDW'01), CEUR Workshop Proceedings, Vol. 39, CEUR WS-org (2001) 6(1-13)

[11] Priebe, T., Pernul G.: Ontology-based Integration of OLAP and Information Retrieval. In: Int. Workshop on Database and Expert Systems Applications (DEXA'03). IEEE Computer Society (2003) 610-614

[12] Pedersen, D., Riis, K., Pedersen, T.B.: XML Extended OLAP Querying. In: Proc. Int. Conf. on Scientific and Statistical Database Management (SSDBM'02). IEEE Computer Society (2002) 195-206

[13] Pedersen D., Riis, K., Pedersen, T.B.: A Powerful and SQL-Compatible Data Model and Query Language for OLAP. In: Database Technologies, Proc. Australasian Database Conference (ADC'02). Australian Computer Society (2002)

[14] Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., Richardson, W.S,: Evidence-Based Medicine: What It Is and What It Isn't. British Medical J., Vol. 312 (7032). BMJ Publishing Group (1996) 71-72

[15] Sheth, A.P., Larson, J.A.: Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. ACM Computing Surveys, Vol. 22 (3). ACM Press, New York (1990) 183-236

[16] World Health Organization (WHO): International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Version for 2003 http://www3.who.int/icd/vol1htm2003/fr-icd.htm (Last access: April 10, 2006)

[17] World Wide Web Consortium: XML Path Language (XPath) v. 1.0. W3C recommendation (as of November 16, 1999). http://www.w3.org/TR/1999/REC-xpath-19991116

[18] World Wide Web Consortium: OWL Web Ontology Language. W3C Recommendation (as of February 10, 2004). http://www.w3.org/TR/2004/REC-owl-features-20040210/

[19] XML Topic Maps (XTM) 1.0, TopicMaps.Org Specification (as of June 6, 2001) http://www.topicmaps.org/xtm/1.0/xtm1-20010806.html

[20] Princeton University Cognitive Science Laboratory: WordNet, a lexical database for English language. http://wordnet.princeton.edu/ (Last access: April 10, 2006)

[21] Health Level Seven (HL7), www.hl7.org (Last access: April 10, 2006)