Roll Out of the OpEN.SC Project

Danilo Schmidt, Gabriela Lindemann – von Trzebiatowski

Humboldt University of Berlin
Department of Computer Science

Lab of Artificial Intelligence

gabriela.lindemann@uv.hu-berlin.de

danilo.schmidt@charité.de

Abstract. This paper is supposed to give a résumé of five years of developing the Open European Nephrology Science Centre – OpEN.SC project. It refers to the last publications about this project and summarizes the itemized phases and components of this development. After the introduction we describe the initial situation and the aims of the OpEN.SC initiative. We give an explanation of the idea of the data structure, the development of an intelligent catalogue and the evaluation of the test results. At last we present screenshots of the OpEN.SC portal.

Keywords: Information retrieval, meta-data repository, medical data, intelligent catalogue

Introduction

The aim of the OpEN.SC project has been the implementation of a metadata repository for clinical data, data of studies, literature and virtual slides. The system works as a data centre to improve the availability of standardized raw data including digitalized glass slides (Virtual Microscopy) and works as an interface for scientific information [LINDEMANN ET. AL. 2006]. Concerning the data of the project partners, their properties and their preferences, the service can promote the creation and organization of a science network with several clinical partners in Berlin and abroad.

The first phase of our project was dedicated to the system design, the database structure, the identification of medical business processes and the interface description of the underlying web services [KEUNE ET.AL. 2007], [LINDEMANN ET.AL. 2007], [SCHRADER ET.AL. 2007].

One special part of the whole project is the Intelligent Catalogue which is implemented as a web service itself. In opposite to standard database requests (for relational databases in SQL) which can only give back cases fitting the query exactly

CONCURRENCY, SPECIFICATION AND PROGRAMMING
M. Szczuka et al. (eds.): Proceedings of the international workshop CS&P 2011
September 28-30, Pułtusk, Poland, pp. 454-464

it gives the possibility to identify similar cases to an inquired one. The Intelligent Catalogue is based on Case Base Reasoning techniques coming from Artificial Intelligence. The underlying data structure is here a Case Retrieval Net – developed by [LENZ ET.AL. 1998] which works together with a spreading activation algorithm. In course of our project the Intelligent Catalogue works now stable in a dynamical version.

The last phase was dedicated to the evaluation of the performance of the catalogue with respect to the achieved results of the retrieval. It is usual in information retrieval settings to take the two measures precision and recall for evaluation.

The aims of OpEN.SC

The establishment of a centre for research information is of crucial importance for the development of new treatment concepts in medicine. The nephrology concerns themselves with illnesses of the kidney including the organ replacement (dialysis, transplantation). Illnesses of the kidney cause a considerable cost in the German health service and worldwide. Considering Germany, so annually more than 2.5 billion Euro is necessary for the financiation of dialysis treatments, whereby the incidence of the dialysis-requiring kidney insufficiency raises annually around 5-8%. The transplantation of a kidney as alternative to dialysis is reduced by the small availability of donor organs.

The Open European Nephrology Science Centre – OpEN.SC - is a co-operation project of the Charité, University Hospital Berlin, the Humboldt-University and Freie University of Berlin and includes in the core the following partners:

- Four Departments of Nephrology (Charité: Campus Mitte, Campus Berlin Buch, Campus Virchow Klinikum and Campus Benjamin Franklin)
- Institute of Pathology, Charité Campus Mitte
- Institute of Medical Informatics
- Artificial Intelligence of the Institute for Computer Science of Humboldt-University.

The topic was the development, maintenance and evaluation of a metadata repository for medical data, virtual slides and other data sources for the domain of nephrology and nephropathology.

The Charité, University Hospital Berlin, has four departments for the treatment of nephrological diseases due to the historical development of the reunion of the clinical centers of the University Hospitals of the Humboldt-University and Freie University of Berlin.

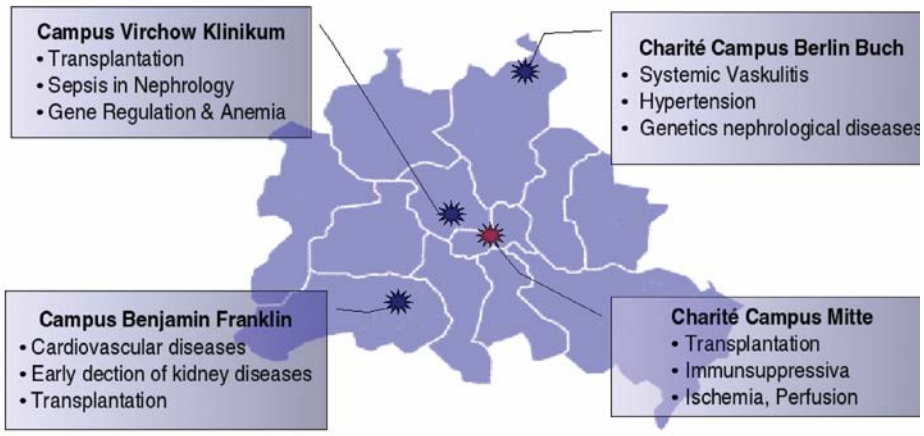

Figure 1: Nephrology Departments of Charité in Berlin

The main idea of OpEN.SC was the implementation of a metadata repository for clinical data, data of studies, literature and virtual slides. The data has been stored and proceed by an intelligent data management tool - the Intelligent Catalogue. The structured data has been specifically presented for different purpose types such as looking for a rare disease or a specific diagnosis or a distinct pattern in the virtual slides of a case. The system works as a data centre to improve the availability of standardized raw data including digitalized glass slides (Virtual Microscopy) and works as an interface for scientific publications

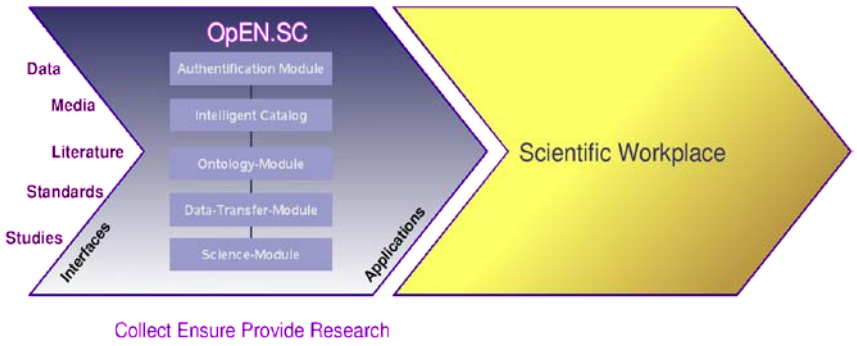

Figure 2: System model of OpEN.SC

Every one of the four mentioned nephrology departments treats 3000 - 4000 cases per year. The availability of all cases of Berlin's university nephrology departments, the sum of about 12.000 cases, would open completely new possibilities of detection of rare disease patterns and quality management.

Concerning the data of the project partners, their properties and their preferences the service can promote the creation and organization of a science network with national and international partners.

RDF Data structure in OpEN.SC

As presented in several papers before [e.g. LINDEMANN ET. AL. 2000, SCHRÖTER ET.AL. 2000] the meanwhile in 17 German transplant centres used web-based Electronic Patient Record database TBase© was our first data source. TBase© is based on a "normal" relational database structure, what means, that it has fixed patterns of entities and their relations, see Figure 3 – only to get an impression. In OpEN.SC we integrate data sources of several involved partner clinics wherein the process of data messaging the private patient data were anonymoused.Every of the clinical partners have their own data formats and structures. Therefore it was important to construct a database model which is independent from the resources. We decided to use the very flexible and open Resource Description Framework (RDF) which leads to a flat database structure consisting only of triples.

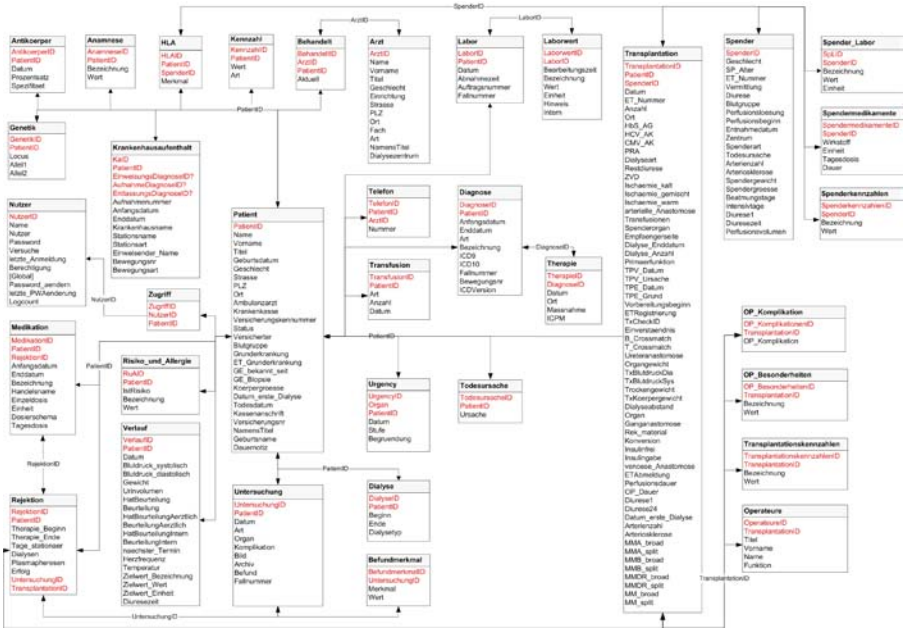

Figure 3: TBase© data structure

RDF - the Resource Description Framework - is a model and language based on XML for representing information about resources of different types. It is particularly useful for storing metadata about shared resources. The design of RDF is generally intended to meet the following goals:

- having a simple data model
- having formal semantics and provable inference
- using an extensible URI (Uniform Resource Identifiers)-based vocabulary
- using an XML-based syntax and supporting XML schema datatypes
- allowing anyone to make statements about any resource

In discussion with our medical partners we divided the complete TBase© domain and additional available data e.g. coming from the VMscope (virtual slides) in seven sub-domains according to considerations about a logical arrangement of medical and further data of patients, see Figure 4. These are:

- Examination Data
- Diagnostic Data
- Treatment Data
- Basic Data (of a patient)
- Administrative Data (e.g. case number in a hospital, SAP data)
- External Data (e.g. virtual slides, discharge letters)
- Project Data (internal)

where the main medical data are in the first three domains.

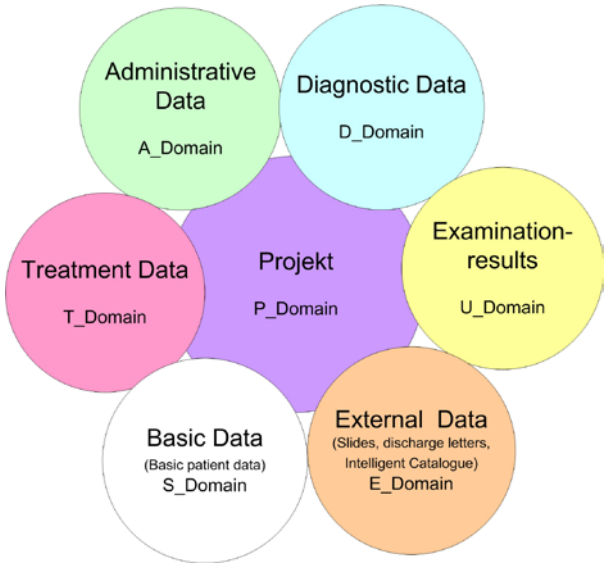

Figure 4: Domains of the OpEN.SC Database

The Intelligent Catalogue

Nowadays it is a well known fact that in the daily routine of hospitals there is produced a huge amount of electronic data coming as output from various diagnostic devices and written parts from physicians and hospital staff. These data have their own different internal formats and are normally stored in more or less isolated databases. Attempts of the last years have gone in the direction to integrate these data in hospital information systems for hospital-wide access for authorized employees. Unfortunately, there exist no standard how to do this independently of a specific

hospital. Therefore large hospitals have now their own information system mostly accessible by Intranet.

Although a physician could internally get “all” patient data he needs there is the problem left how to filter and present only the necessary ones. Moreover, he has no access to even anonymous data of other sources if he needs such for evaluate a diagnosis, look for appropriate medication, a seldom case, self-education or to realize a medical study.

Exactly for these purposes we have developed an Intelligent Catalogue at first in the areas of Nephrology and Pathology. As a second core besides the Virtual Microscopy we used the web-based Electronic Patient Record TBase© which was implemented in a German kidney transplantation program as cooperation between the Nephrology of Charité Campus Mitte and Virchow and the AI Lab of the Institute of Computer Sciences of Humboldt University of Berlin [SCHRÖTER et. al. 2000], [LINDEMANN ET. AL. 2000]. At present it comprises records of nearly 9754 transplant recipients and patients on the waiting list. Currently TBase© automatically integrates essential data from the laboratory, clinical pharmacology, nuclear medicine, findings from radiology and administrative data from the SAP-system of the Charité. Furthermore it is in the daily routine at 17 transplantation centers.

The main idea of the Intelligent Catalogue is to collect, combine and structure all data with respect to the medical domains of Nephrology and Pathology and to allow a quick and easy access to these data for the above mentioned purposes. The mean to do this is the technique of Cased-Based Reasoning (CBR) coming from Artificial Intelligence. In contrast to classical database inquiries where users often have to formulate strictly syntactical database-queries – and physicians are mainly not familiar with this – modern CBR-systems allow to formulate questions in natural language. Moreover, the underlying knowledge base of the CBR-systems is not organized like a relational data-base but as a form of an associative memory - the Case-Retrieval Nets (CRN) [HILDEBRAND ET. AL. 2009].

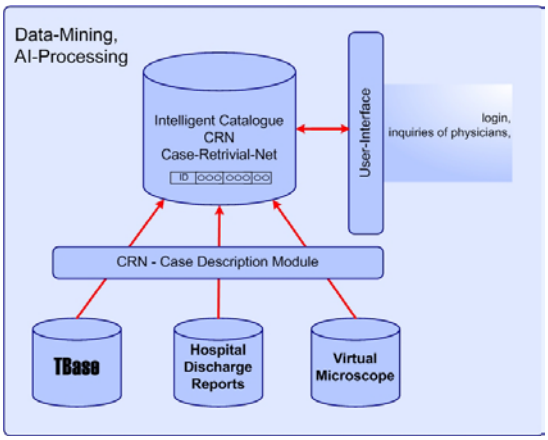

Figure 5: Intelligent Catalogue

Case Retrieval Net in OpEN.SC

A Case Retrieval Net (CRN) is structured as a net of information entities, in which the cases are represented as partial graphs of this net with weighted arcs. The net consists of nodes for the identification of each case (descriptor) and nodes for the representation of the information entities (IE's) which characterize a certain case. Each case descriptor is connected with its IE nodes via relevance arcs, where the weights of arcs intend to reflect the relevance of each kind of IE or attribute, respectively. In addition there are weights of arcs between IE nodes of a same attribute in different cases. They indicate the value of the local similarity function for two IE's and are computed in the retrieval process (figure 6).

A suitable retrieval method is a Spreading Activation algorithm. The inquiry activates IE nodes, which spreads its activation then over the similarity and relevance arcs. This leads to a half-order in the descriptors, which is induced by the similarity of each case to the inquiry. Thus CRN's are particularly useful on the one hand for the administration of large quantities of case data and on the other hand for the fast access to these data. The cases must not be scanned sequentially during the retrieval process, but can be regained in the net structure efficiently. The structure of a CRN allows to add new cases and therefore new information units (e.g. with the establishment of new medical diagnostic procedures) with no problems [HILDEBRAND ET. AL. 2008].

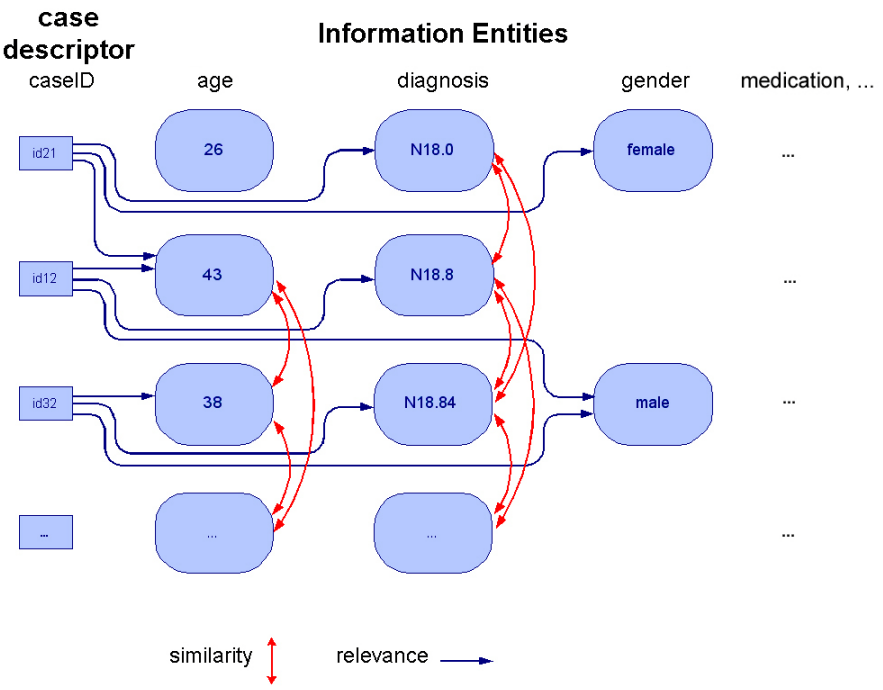

Figure 6: Case Retrieval Net

Evaluation of the Intelligent Catalogue

One part of the evaluation of the whole OpEN.SC system was to find out, how many located or missing results are generated by a query to the Intelligent Catalogue via the OpEN.SC portal which processes the retrieval of the data of the OpEN.SC metadata repository.

The repository includes different kinds of medical data of 9754 patients. To evaluate the retrieval which was developed by us, we used three several datasets for testing. These datasets were randomly chosen out of all patient records and comprise data of 50 patient cases each including about 1300 diagnoses. Each patient has 24 diagnoses on average. For the evaluation we have created 10 queries for the search mask of the OpEN.SC portal. In order that a sufficient number of results are generated by the retrieval algorithm, we have designed the queries in this way, that they include diagnoses which are common diagnosed in the Nephrology domain.

Our results show that with respect to the recall most of them have an outcome of 100 percent. See figure 7, as an example. For the other results please have a look in the paper [SCHMIDT ET. AL. 2010]. The retrieval results with less than 100 percent found cases have as maximum one or two missing cases. The reason for these missing results are cases that are not correctly coded by the WHO ICD-10 code in the source databases. That has historical reasons because some of the physicians did not do that. Figure 7 and Table 1 demonstrate that there is a high success rate to find cases in the data repository by use of our Intelligent Catalogue.

Evaluation Dataset 1

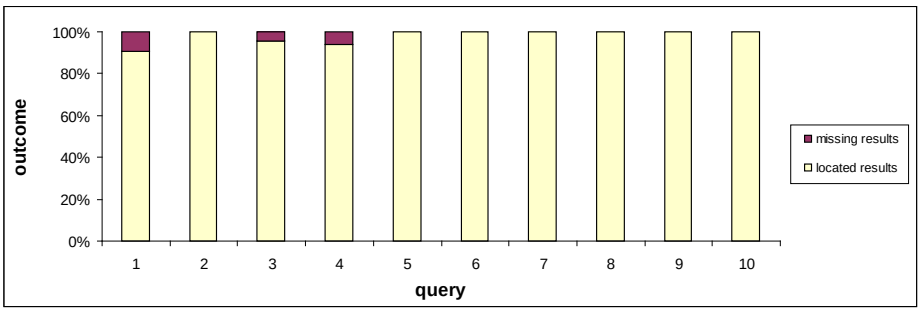

Figure 7: results of the evaluation for dataset 1

query	1	2	3	4	5	6	7	8	9	10
located results	19	9	22	32	10	11	12	35	37	6
missing results	2	0	1	2	0	0	0	0	0	0

Table 1: results of the evaluation for dataset 1

Conclusion

The Open European Nephrology Science Centre - OpEN.SC was born out of the idea to develop a metadata repository for clinical data, data of studies, literature and virtual slides. The first phase of the project was dedicated to the system design, the database structure, the identification of medical business processes and the interface description of the underlying web services.

One special part of the whole project is the Intelligent Catalogue which is implemented as a web service itself. In opposite to standard database requests (for relational databases in SQL) which can only give back cases fitting the query exactly it gives the possibility to identify similar cases to an inquired one. The Intelligent Catalogue is based on Case Base Reasoning techniques coming from Artificial Intelligence. During the project phase we realized that the transformation of our data structure from a static to a dynamical one leads to a reduction of memory space as well – and this was unexpected before – as to computational time. Moreover we introduced some more attributes in the next version of the whole OpEN.SC project, like medication and information about transplantation data. By the use of the dynamical data structure for the Case Retrieval Net the extension of the net was very easy by only adding information entities for the new kinds of values. In the following figure 8 and figure 9 one get an impression of the look of the OpEN.SC portal site for queries of clinical users including the possibility to ask for medication and transplantation data.

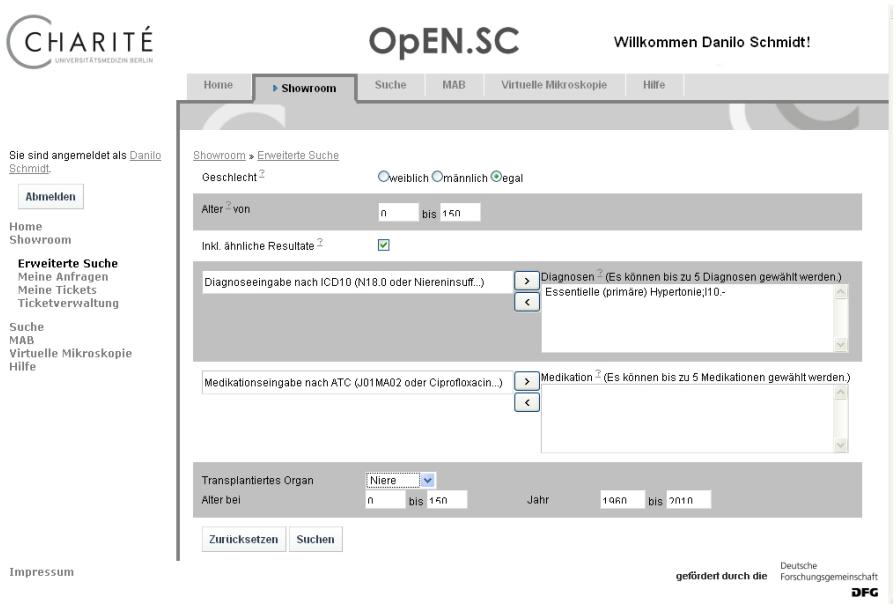

Figure 8: Snapshot of the OpEN.SC Portal for a query

The last phase was dedicated to the evaluation of the performance of the catalogue with respect to the achieved results of the retrieval. The results show that the

Intelligent Catalogue works well and that with respect to the recall the results have an outcome of almost 100 percent. Since the beginning of 2011 the OpEN.SC portal has been onlined for authorized physician and supports them as a scientific workplace.

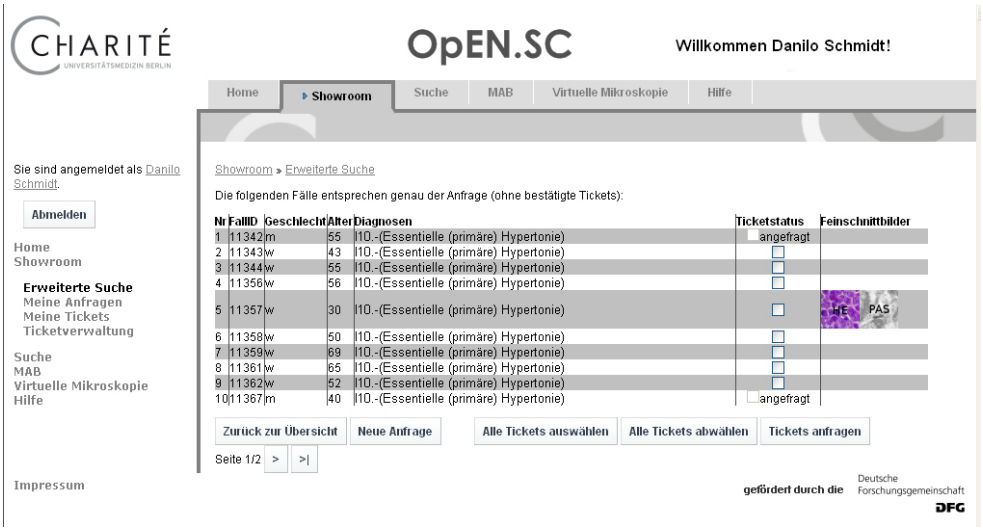

Figure 9: Snapshot of the result of a query

Acknowledgement

This work is partly granted by the Deutsche Forschungsgemeinschaft (DFG) under the main point program “Centres of Research Information”. OpEN.SC is a collaboration of the Department of Computer Science - Artificial Intelligence, the Institute of Pathology, the Institute of Medical Informatics and the four Departments of Nephrology of the Charité in Berlin as well as associated partners from abroad.

References:

[HILDEBRAND ET. AL. 2008] F. Hildebrand, G. Lindemann, D. Schmidt: Application of Case Retrieval Nets in the OpEN.SC Project, Informatikberichte, Concurrency, Specification and Programming CS&P'2008, pp. 218-228, ISSN: 0863-095X, Groß Väter See, 29. September – 01. Oktober, Germany, 2008.

[HILDEBRAND ET. AL. 2009] F. Hildebrand, G. Lindemann, D. Schmidt: A Dynamical Data-Structure for Case Retrieval Nets in the OpEN.SC Project, ICICIS 2009 Workshop "Medical Informatics and e-Health", Cairo 2009

[KEUNE ET. AL. 2007] D. Keune, G. Lindemann, D. Schmidt, T. Schrader: A modern Database Structure for Patient Data within the OpEN.SC project, Proceedings of the CS&P2007, Concurrency, Specification & Programming 2007, pp. 357-364, ISBN: 978-83-88374-28-9, Lagow, 27.-29. September, Poland, 2007.

[LENZ ET. AL. 1998] M. Lenz, B. Bartsch-Spörl, H.-D. Burkhard, S. Wess (Eds.): Case-Based Reasoning Technology, Springer, 1998.

[LINDEMANN ET. AL. 2000] G. Lindemann, L. Fritsche: Web-Based Patient Records - The Design of TBase2. In Bruch, Köckerling, Bouchard, Schug-Paß (Eds.) New Aspects of High Technology in Medicine; Seiten 409-414; 2000.

[LINDEMANN ET. AL. 2006] G. Lindemann, T. Schrader, D. Schmidt: The Aim of the Open European Nephrology Science Center (OpEN.SC) – First Steps. In: Proceedings of "Concurrency, Specification & Programming" – CS&P2006; Vol. I, II, III. Informatik Berichte, Institute of Computer Science, Humboldt University Berlin, ISSN: 0863-95X, Germany, 2006.

[LINDEMANN ET. AL. 2007] G. Lindemann, D. Schmidt, T. Schrader, D. Keune: The Resource Description Framework (RDF) as a modern Structure for Medical Data, Proceedings of World Academy of Science, Engineering and Technology, CESSE 2007, pp. 422-427, ISSN 1307-6884, Venice, 23.-25. November, Italy, 2007.

[SCHMIDT ET. AL. 2010] D. Schmidt , G. Lindemann,: Precision and Recall of the Intelligent Catalogue in the OpEN.SC - Extended Abstract -. In: Proceedings of "Concurrency, Specification & Programming" – CS&P2010; Vol. I, II, III. Informatik Berichte, Institute of Computer Science, Humboldt University Berlin, ISSN: 0863-95X, Germany, 2010.

[SCHRADER ET. AL. 2007] T. Schrader, D. Schmidt, G. Lindemann, S. Niepage, C. Hahn, D. Keune, T. Wetzel, S. Krieger, T. Weckend, H. Peters, K. Budde, B. Rudolf: The Open European Nephrology Science Center: An International Research Resource Centre for Kidney Diseases, Iran Journal of Kidney Diseases, 11th International Congress of Nephrology, Dialysis and Transplantation, pp. 4-5, ISSN 1735-8582, 21.-23. November, Iran, Tehran 2007.

[SCHRÖTER ET.AL. 2000] K. Schröter, G. Lindemann, L. Fritsche: TBase2 – A web-based Electronic Patient Record. Fundamenta Informaticae 43, 343-353, IOS Press, Amsterdam, 2000