IKTA5-126/02.

Summary in English

LOBO: LOgic-Based management of Ontologies

The proposed project aims at developing a logic-based ontology management tool, based on the decade long national research activities and the results of successful European research projects of the participants. The targeted LOBO (LOgic-Based management of Ontologies) workbench is a domain independent tool for managing ontologies. However, the LOBO pilot application in the biomedical domain is an extremely important part of the project, as well. Today, biomedicine is one of the largest supplier and user of ontologies. Most of the experiences to be gained, as well as the solutions to be created in this domain are expected to be easily transported into other domains.

The project proposal builds on the results of the SILK (System Integration via Logic and Knowledge) EU-project, No. IST-1999-11135, coordinated by one of the consortium members. Within this project, a **base technology** has been developed for the support of information integration at application level. This technology is intended to be further developed by the current proposal in the direction of conceptual level, i.e. ontology based, integration. Following the technical direction of the related EU-project, the project proposal regards the use of, and conformance to, the relevant and practical IT standards (such as Unified Modelling Language, Object Constraint Language, Prolog, Java, eXtended Markup Language, XML Model Inter-change) as a fundamental requirement for the technology development. Since the proposed project takes the technology, created in the above-mentioned EU-project, as a starting point, the feasibility risk of the proposal is substantially decreased.

However, the further development of the base technology is only feasible if performed in a real application domain context. The consortium plans to meet this requirement by applying the technology in the biomedical domain for reasons detailed below.

Following the rapid developments in the 90's, the biological medicine (and the related fields of molecular biology, genetic engineering etc.) produces an exponentially increasing **volume of biological data**. Beside the sheer volume of data, the underlying biological **complexity** makes it an even bigger challenge to handle this type of information, and this goes together with the problem of **distributedness**, e.g. several hundreds of databases of genomic information are maintained by experts all over the world, and the situation is similar in other areas. The "usual" solution of creating one large central database, or data warehouse, of these information sources is completely out of question, due to the distributedness and to the rapid growth of information and knowledge. On the other hand, biologists have a pressing need for support, guidance, for the analysis of their data, and for the collaboration with other researchers. For the better integration of biological and medical sciences, for the interconnection of information and knowledge in different domains as well as for their application in a unified framework, innovative use of information technology is required through a practical, **common platform**.

The biomedical ontologies play a significant role within the actual set of tools of the problem area. These ontologies codify the explicit body of knowledge in a specific domain, and provide it in a structured form for the use of domain experts. By now, ontologies are increasingly used as necessary utility tools in the course of biomedical research, experiments and even in everyday practice.

However, there are a number of difficulties which hinder the spreading of ontologies and their multiple use: 1) heterogeneous, non-standard storage mechanisms, 2) incomplete and/or not

sufficient support for individual experts, 3) differing viewpoints and levels of domain descriptions which hinder further modes of usage. Other factors, like handling the non-usual sizes of biomedical ontologies, ensuring their consistency and reusability, keeping them language independent, create just additional problems.

The proposed ontology development technology and tools are going to provide new solutions primarily in the field of **consistency control** of ontologies, **interconnection** and integration of domain ontologies with expert- or expert group-specific ontologies, and ontology based (i.e. **conceptual level**) **inquiries** of domain information sources. This core set of functionality is extend by other noteworthy features, like support for automatic **restructuring** of ontologies (under human supervision) which promises traceability of rapid changes, and intuitive graphical **presentation** of the structure and content of ontologies as well as the retrieved information.

The result of the project is not only a usable, general purpose (domain independent) tool but it will represent an important contribution to the IT toolset of the experts of domestic biomedicine. With the help of the tool, local expert groups can easily organise their own conceptual view in the form of ontology, and compare this with other more "official" ontologies for filtering out deficiencies, errors and/or inconsistencies. This could accelerate technical cooperation between different expert and research groups when naming and structuring newly emerging concepts. This could also represent a substantial improvement in the ability to use and interpret biomedical data produced at an ever-increasing speed.

The technological basis of the LOBO-workbench (**Prolog** based knowledge storage mechanism and related programming interface, and **Java** based data access and user interfaces) allows for relatively easy enhancement of ontologies with constraint logic based, domain specific inference and knowledge management modules. This expands further the borders of IT support for research activities of expert groups.

The results of the project will be verified through two experimental applications. These involve the creation of a data/knowledge base containing surgical events, and a data/knowledge base supporting the evaluation of the mammographical exposures, both within the institution of the medical partner.

The No 1. consortium member is a leading company of the domestic software market, specialised in application development and integration. It has also a notable R&D section, which proved its abilities in a dozen of national and European projects in the past decade. In the proposed project, this consortium member is going to play the role of **technology centre**.

The No 2. consortium member has several years of experience and practice in applying ontologies in different domains, and in biomedicine, in particular. In the proposed project, this consortium member is going to play the role of **knowledge centre**.

The No 3. consortium member represents the medical field, and thus plays the role of the biomedical **knowledge centre**, as well as that of the **application centre**. This consortium member will play a crucial role in pilot application development and evaluation.