Proposal for a research cooperation between:
CNR IP, Rome, Italy and Iwate Prefectural University, Morioka, Japan

An ontological analysis of user requirements in the context of Lyee software engineering methodology

Introduction
Lyee is a powerful and innovative software engineering methodology and tool for implementing software. It has been shown that such a tool allows an implementation faster than by using traditional methods.

Lyee has also a revolutionary approach in the requirement analysis phase of the software life cycle: the user's requirement can be entered directly in the system without the usual preliminary mediation of its formalization.

Users express their requirements by means of words and required screen input and output forms. Some time is spent by Lyee designers in order to select the useful information among that provided by users.

The ontological approach
In order to better understand how the ontological approach fits into the Lyee framework and how it could improve its performances, it is worthwhile to sketch out the meaning of ontology.

Apart from its definition in a philosophical context - where it refers to the subject of existence - ontology in our context is "a partial specification of a conceptualization". Recently Sowa proposed the following definition influenced by Leibniz:

“The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D. [...]”

Some admit informal descriptions and hierarchies, only aimed at organizing some uses of natural language; others require that an ontology be a theory, i.e. a formal vocabulary with axioms defined on such vocabulary, possibly with the help of some axiom schema, as in description logics.
In our perspective, an ontology is a formal theory that partially specifies the conceptualization (i.e. the intended meaning) of a lexical item as it is used in a certain domain. Since lexical items are often used with more than one conceptualization in the same domain (they are "polysemous"), such different conceptualizations have to be specified and segregated within different formal contexts, or conceptualizations must have assigned distinct names within the same context. A “context” is a theory that serves as a module within a system that allows a partial ordering among its component theories. The procedure by which the lexical items are conceptually analyzed and their conceptualizations are (partially) specified within a context hierarchy is what we call the "ontological analysis".
Goals

The aim of this research is to provide an ontological analysis of the possible user requirements in order to understand which type of requirements fit in the Lyee framework.

Currently, some time is spent by Lyee designers in order to single out only those words which are admitted as valid requirement words. In fact, all the other words are discarded and their meaning subsequently recovered.

Lyee makes some explicit commitment to the data types it takes into account during requirement analysis:
• "Words" (considered the only essential to design a program): this is actually conceptualized as a lexical item that conveys information about commands to be implemented in the program. The tentative grammatical guessing by Lyee designers is that only some "nouns" (object nominals, first-level noun phrases) convey such information, but evidence from actual Lyee requirement analyses shows the presence of other kinds of grammatical categories.
• Screens (desiderata about and sketches of the screens that should be generated by the final program).
• Files (specification of the files to be included in or generated by the program).
• Predicate structures (input/output specifications of each requirement element extracted from "words").
• Routes (sequences of commands required).

On this basis, CNR aims at experimenting the development of an ontology of requirements as a framework for Lyee to perform a more focused and quicker requirement analysis, even without entering the time-consuming and user-alienating modelling of classical requirement analysis.

For example, a preliminary taxonomy of Requirement Documents (RD) is the following:
• According to the semiotic type, a RD can be a text, a picture, a conversation, a drawing, a printout.
• According to the level of formality or structuring,
  • A text can be free, structured in some way, or formalized
  • A drawing can be informal, diagrammatic, or formalized
• According to components,
  • A text can contain linguistic items, concepts and relations, etc.
  • A drawing can contain nodes and links, formal boxes of various sorts, etc.
• According to its functional content, a RD can contain descriptive, procedural, connective, or meta-level information.
• According to its topic, a RD contains information relevant for some domain, task, participant type.
• According to its expected outputs, an RD can contain expectations about: given values, calculated values, relations between components, files, screens and windows, diagrammatic representations, pictures, etc.
• According to its linguistic form, a RD can contain explicit or implicit information. After being explicated, implicit information should be classified according to this taxonomy as well (iteration).

An ontological analysis of the admissible requirement words can help Lyee designers in extracting knowledge by the users. Having a model of admissible and not admissible word, the Lyee designers will be able to select the right words in a faster way.

In such a way, an expected benefit of such an approach is to make a life-cycle of a project even shorter, by cutting the time needed in the first interaction with users.

Workplan
Year 1
Investigation on Lyee’s requirement analysis
Preliminary taxonomy of requirement documents
Ontology of user requirements (first version)

Year 2
Ontology of user requirements (revised version)
Analysis of existing term extraction tools
Choice of a tool for term extraction and ontological encoding of terms
Preliminary design of a tool for Ontological Requirement Analysis (TORA)

Year 3
Design and implementation of a tool for Ontological Requirement Analysis (TORA)
Design and implementation of an interface between TORA and Lyee

N.B. Implementation of tools will be made in co-operation with Iwate University and supported by Iwate human resources, in Italy and/or in Japan

People and working environment

The “Consiglio Nazionale delle Ricerche” (CNR) (National Research Council) is the technical body for basic and applied research of the Italian Ministry of University and Research. It is also in charge of the preparation of standards of national interest and manages special pluri-annual grants for "Finalized Projects" on focused topics, following the policy directions of the Council of Ministries. It is organized in about 200 autonomous institutes and centers in all the branches of sciences and humanities.

The research activity of the Ontologies and Conceptual Modeling Group of CNR-IP is centered on the semantic and ontological representation of medical terms and clinical guidelines. Several prototypes have been implemented for the management of terminologies, for the design and management of multimedia patient folders (the CADMIO multimedia patient folder) and for supporting clinical guidelines management (the SMART system).

The group took part - among the others - in the following projects: the CNR projects: SOLMC (Ontological and Linguistic Tools for Conceptual Modelling) and ONTOINT (Ontological Tools for Information Integration) and the EU projects: GALEN, HANSA (on the integration of heterogeneous clinical information systems) and PROGUIDE (on the development and dissemination of clinical guidelines).

Current research activities are:
- Development of methodologies for the integration of heterogeneous information.
- Modelling of ontologies for medical information integration.
- Formal modelling of clinical guidelines.
- Tools for clinical guidelines authoring.
- Description logics for conceptual modelling in medicine.
- Development of tools for cooperative modelling over the WWW.
- Ontology-based access to distributed information.
Current products include:
- The ONIONS (Ontological Integration On Naïve Sources) methodology for the alignment of conceptually heterogeneous terminologies.
- The ON9 ontology library, including about fifty modules describing both generic and domain ontologies (about 1,500 concepts, 3,000 axioms). ON9 provides the framework to integrate the top-levels of five authoritative medical terminologies (UMLS, SNOMED-III, ICD-10, Gabrieli, GALEN).
- The representation of the terminological data of the UMLS (Unified Medical Language System) project in the description logic-based Loom knowledge representation system and its (ongoing) ontologization. The current database features the subsumption of the about 470,000 UMLS concepts under some ON9 concept, the representation classification of more than 310,000 concepts from within the repository, and the formal translation of about 50,000 templates stated in the UMLS. Such investigation also provides a solution for the polysemous multi-classification of UMLS concepts.

Domenico M. Pisanelli. (1959) graduated in Electronic Engineering (Computer Science in Medicine) at the University of Rome in 1986. Since 1987 he is researcher at ITBM-CNR (National Research Council, Institute of Biomedical Technologies). His main research interest is in the field of medical informatics (expert systems in medicine, computerized guidelines, and telemedicine). From 1990 he held an appointment at University of L’Aquila for teaching computer science in the Engineering Faculty. He is involved in several national and international projects on computer science in medicine and in European working groups on standardization in health care. He is author of the book "Il medico in rete" ("The Physician on-line") published by Laterza and of over 80 scientific papers (published in international journals and conferences) in the areas of statistical databases and medical informatics. He is member of the IEEE Computer Society and Engineering in Medicine and Biology Society and of the Association of Computing Machinery.

Aldo Gangemi. (1962) graduated in Philosophy (Analysis of Scientific Languages) at the University of Rome in 1989. Since 1995 he is researcher at ITBM-CNR (National Research Council, Institute of Biomedical Technologies). His main research interest is in the fields of formal ontologies and methodologies for information integration. He has been involved in several national and international projects on medical informatics and ontological tools for conceptual modeling. He is author of about 40 scientific papers (published in international journals and conferences) in the areas of artificial intelligence, medical informatics, linguistics, and philosophy.

Gerardo Steve. (1941). graduated in Mathematics at the University of Rome. Currently he is Senior Researcher at ITBM-CNR (National Research Council, Institute of Biomedical Technologies). He worked in the field of probabilistic reasoning and treatment of uncertainty also dealing with application to medical expert systems. His main research interest is currently formal ontology. He has been involved in several national and international projects on medical informatics and ontological tools for conceptual modeling. He is author of several scientific papers (published in international journals and conferences) on computer science and AI in medicine.