Socio-technical Ontology Development for Modelling Sensemaking in Heterogeneous Domains

Dhavalkumar Thakker, Fan Yang-Turner, Lydia Lau, Vania Dimitrova

School of Computing, University of Leeds Leeds LS2 9JT, United Kingdom {D.Thakker, F.Yang-Turner, L.M.S.Lau, V.G.Dimitrova}@ leeds.ac.uk

Abstract. Sensemaking is often associated with processing large or complex amount of data obtained from diverse and distributed sources. With information explosion from the web, sensemaking is becoming ubiquitous and ever more challenging. Semantic technologies have potential to support understanding of sensemaking process with the benefits they bring (e.g. reasoning, aggregation, automation). Conceptual models of sensemaking have been developed to understand its complex processes by social and information scientist. However, these frameworks are not applicable directly to system design. This paper describes a socio-technical approach for modelling sensemaking process in order to inform the development of intelligent services to aid sensemakers. We apply an *a priori* ontology modularisation methodology for handling complexity of heterogeneous domains and utilise well-known sensemaking theoretical framework to guide ontology development. This approach is applied in an EU project - Dicode, for the development of its sensemaking ontology.

Keywords: Ontology, Sensemaking, Modularisation, Semantic Annotation

1 Introduction

Semantic technologies, underpinned by ontologies, have been seen as one of the promising platforms for developing knowledge management systems [1, 2]. Examples of successful ontology developments can be found in a diverse range of domains such as multimedia [3] and life sciences [4, 5]. In these relatively well-defined and well-researched domains, ontological representations enhance the machine's reasoning capability on those knowledge bases.

With the recent proven successes of semantic web and ontologies, the field is ready to take on challenges offered by complex social-oriented domains which are less welldefined or scoped. Sensemaking is such a domain that involves cognitively-complex processes carried out by human and often requires injection of tacit knowledge. Moreover, sensemaking encompasses a range of behaviour surrounding the collection and organisation of information, may be across domains, for better understanding of a situation. Therefore, it is very challenging to derive a systematic and thorough understanding of the sensemaking processes from domain experts using traditional knowledge elicitation techniques. Conceptual models of sensemaking have been developed to understand its complex processes by social and information scientists [6, 7, 8]. However, the problem of understanding and supporting sensemaking via technology remains challenging [9]. Initial work has already started in utilising semantic technologies for aiding sensemaking process in the domains of linked data [10], visualisation [11] and e-health [12]; which focus on applications that serve sensemaking rather than modelling sensemaking as a generic process.

This paper proposes a socio-technical approach for the development of an ontology which models sensemaking process in order to inform the design of intelligent aids to sensemakers. This is motivated by the vision of an EU project - Dicode¹, which aims to provide synergy between human and machine intelligence in collaboration and decision making within data-intensive environments. Theoretical frameworks on sensemaking, combined with an *a priori* ontology modularisation methodology, are used to guide the ontology development for sensemaking in heterogonous domains.

The paper is organised as follows. Section 2 explores the domain of sensemaking and the issues to be considered in developing a sensemaking ontology. Our sociotechnical approach is proposed in section 3. Section 4 illustrates the application of the socio-technical approach in Dicode for the development of a multi-layered **Dicode ONtology** (referred as **DON** afterwards) for sensemaking. To better understand the potential benefits of semantics (e.g. using the ontology for reasoning, aggregation, automation) for applications in sensemaking domains, a proof-of-concept prototype "*Augmentor*" has been developed and discussed in Section 5. The concluding section summarises our contribution and future work.

2 Sensemaking: A Case Study

Sensemaking, as in "to make sense", is a process of transforming information into a knowledge product [8]. Sensemaking process involves interplay between foraging for information and abstracting the information into a representation called a schema that will facilitate a decision or solution [6]. It is often associated with processing large or complex amount of data obtained from diverse and distributed sources. There has been a recent increase of interest in sensemaking driven by the information explosion from the web that has rapidly changed our ability to assess large amounts of information [20].

Dicode project is aimed at supporting sensemaking and decision making in dataintensive and cognitively-complex settings. The solution foreseen in the Dicode project will bring together the reasoning capabilities of both the machine and the human. There are three use case partners involved to validate the transferability of Dicode solutions. They are from three different domains: (1) **Clinico-Genomic (CG)** research where clinical research professionals collaborate to explore scientific findings related to breast cancer using very large datasets; (2) **Rheumatoid Arthritis (RA)** clinical trial where medical personnel involved in the clinical trials collaborate and exchange their professional judgment within complex clinical decision making

¹ <u>http://www.dicode-project.eu/</u>

processes; and (3) **Public Opinion (PO)** monitoring where analysts watch social media to monitor public perceptions of their clients' branding, products or services.

These three use case partners were selected to address common challenges in sensemaking and decision making. All use cases experience the problem of information overload; all require sensemaking towards decision making based on cognitively intensive analysis and interpretation of data; all need to discuss and share interpretation and decision making rationale between specialists. They cover the full range of features and functionalities to be addressed by the project, from various sectors and domains and draw relevant information from large scale and real time data residing in heterogeneous sources. However, beyond these high level similarities, each use case comes from different domains (e.g. biomedical, medical, or public relations), deals with different type of data (e.g. structured database tables, semi-structured log data, unstructured blogs, forum discussions or tweets) from different data sources (e.g. biomedical analysis tools, image analysis software or social media monitoring tools) and with different work practices (e.g. organisational practices of research teams or market research teams for public opinion monitoring).

Both the similarities and the differences among the use cases bring forth several research challenges in terms of ontology development: (1) **Domain complexity**: Understanding sensemaking in these domains is difficult as it involves heterogeneous sources of knowledge, i.e. expertise from multiple disciplines. (2) **Knowledge scope expansion**: The conceptualisation process is generally dynamic and evolves with the increasing amount of tacit knowledge being made explicit. This means that certain concepts and relationships are unidentified in the beginning. Hence, it is not always possible to build an all-encompassing ontology in the very first instance. (3) **Systematic development**: Traditional knowledge elicitation techniques for conceptualisation that rely on domain experts are not sufficient as conceptualisation might result in ad-hoc modelling. To address these challenges in Dicode and for the development of DON, we apply a novel socio-technical sensemaking modelling approach presented in the next section.

3 The Proposed Socio-technical Approach

Socio-technical principles started in the age of shop floor automation [13]. They have since been applied to the design and implementation of computer-based systems and information technology [14, 15]. Underpinning our proposed socio-technical approach for modelling ontology for sensemaking is the concept of *a priori* modularisation. We have developed *a priori* modularisation methodology [16] that enables dividing the domain ontology into several modules from the outset in order to handle the complexity and dynamicity of ontology modelling in ill-defined domains. This modularisation methodology is devised for a class of problems that involve cognitively-complex processes carried out by humans and require tacit knowledge (e.g. decision making, sensemaking). The understanding of such domains involves inter-disciplinary domain experts who often utilise a theoretical framework to guide the articulation of their understanding. According to this methodology, ontology development begins with some *theoretical framework* and arrives to *case specific*

domain ontologies. We follow the three-layered development of domain ontologies which consist of an upper abstract layer, a middle reusable layers, and a lower case specific layer. Each layer may consist of one or more ontology modules (see Fig. 1).

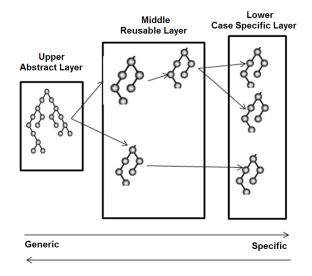


Fig. 1. A Multi-layered Ontology Development with a Priori Modularisation

Upper abstract layer: The chosen socio-driven theoretical framework(s) will have the most influence on this layer when the base concepts for the domain are defined following the theoretical framework. Conceptualisation at this level is conceived and developed *independently from its usage context* and avoids defining any concepts that are tied to a particular use case. The sensemaking theoretical frameworks selected for this approach are discussed in section 4.

Middle reusable layers: Middle layers, which *evolve organically through use*, are used to make the connection between the upper ontology layer and the case specific ontology layer. The concepts captured in this layer are likely to be expanded as more tacit knowledge used for interpreting the base concepts is being captured. This layer provides a context-rich bridge between the upper level concepts and the multiple case specific domain ontologies. The middle layers can expand into a number of sub-layers depending on the commonalities among specific cases. The concepts defined in the sub-level should be reusable and remain high level. Only thinking in terms of reusability [17] will keep this layer generic for any sensemaking domain.

Case specific layer: This layer defines the concepts that are specific to each use case (i.e. closer to the content and usage). During this stage, when commonalities in the use cases are discovered, those ontological statements will be moved to the middle layers. This may lead to the expansion of a module or even start a completely new module in the middle layers.

4. Dicode ONtology (DON) for Sensemaking

The following subsections present the main features of the three-layer **Dicode ONtology** (**DON**) developed for sensemaking.

4.1 Upper Abstract Layer

This upper abstract layer ontology covers base concepts that describe sensemaking process for Dicode. We here explain our choice of upper abstract layer sensemaking frameworks.

In Dicode, each use case involves group of professionals collaborating to address complex problems by combining experience and expertise towards a shared understanding. Hence, *collaborative sensemaking* is the ultimate target for our work.

We are inspired by the work of Paul and Reddy [18] on collaborative sensemaking. Their framework shows collaborative sensemaking activities are often initially split into tasks/sub-tasks and sub-tasks are performed by different group members (possibly by performing individual sensemaking), depending on their roles and expertise. Roles can be organizational or might be assigned informally. It also defines the collaboration triggers (e.g. ambiguity of information, role-based distribution of information, and lack of expertise) and characteristics of collaborative sensemaking (e.g. prioritizing relevant information, sensemaking trajectories, and activity awareness). The framework also highlights the need to bring together individual sensemaking activities prior to supporting the collaborative sensemaking activities.

To address *individual sensemaking*, we adopted a notional model developed by Pirolli and Card [8], in which sensemaking process is defined as two interconnected loops: foraging loop and sensemaking loop. The foraging loop involves sensemaking operations such as searching and filtering information, gradually leading to the identification and organization of relevant knowledge. The sensemaking loop is an iterative development of a mental model from the schema/representation that best fits the evidence, which involves searching for support (e.g. using support systems) and using that schema to complete a final task.

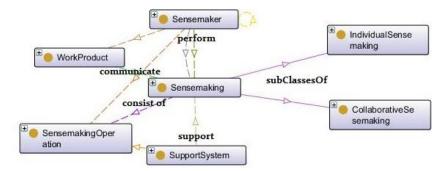


Fig. 2. Abstract Sensemaking Model

Fig. 2 and Table 1 outline the resulting high level base concepts drawn from these two frameworks. The upper abstract layer caters for the main elements of the individual sensemaking such as: actors (e.g. SENSEMAKERS), outcomes (WORK PRODUCT such as documents, diagrams), support services (SUPPORT SYSTEMS, such as data mining, semantic search), and sensemaking operations performed as part of TASKS by human or by machines and main axioms linking them.

Table 1. Abstract Sensemaking Model

Description Logic (DL) syntax
CollaborativeSensemaking ⊑ Sensemaking
IndividualSensemaking 🗆 Sensemaking
Sensemaking $\sqsubseteq \exists$ consistOf.SensemakingOperation
SensemakingOperation $\sqsubseteq \exists$ using. InformationSource
SensemakingOperation $\sqsubseteq \exists$ on. Data
WorkProduct ⊑ ∃ communicate. Sensemaking
Representation \sqsubseteq WorkProduct
Representation $\sqsubseteq \exists$ represent. Data
Sensemaker ⊑ ∃ perform. Sensemaking
Sensemaker $\sqsubseteq \exists$ carry out. SensemakingOperation
Sensemaker ⊑ ∃ have. Expertise
Sensemaker ⊑ ∃ create. WorkProduct
Sensemaker $\sqsubseteq \exists$ utilise. InformationSource
SupportSystem $\sqsubseteq \exists$ support. Sensemaking
SupportSystem $\sqsubseteq \exists$ facilitate. SensemakingOperation
CollaborationTrigger ⊑ ∃ trigger. Collaborativ Sensemaking
Sensemaker $\sqsubseteq \exists$ interactsWith. Sensemaker

The upper abstract layer also contains conceptualisation of collaborative sensemaking process: TRIGGERS triggering COLLABORATIVE SENSEMAKING, SENSEMAKERS interacting with other SENSEMAKERS and playing a ROLE and offering EXPERTISE, division of tasks into SHARED TASK and outcomes (SHARED UNDERSTANDING, SHARED REPRESENTATION).

This upper abstract layer is a starting point for extending into more specific ontologies.

4.2 Middle Reusable Layers

In the middle reusable layers, we defined concepts and respective modules that are used across three use cases. The middle layers in the sensemaking ontology include common concepts that expand the base concepts from the upper layer. The common concepts within all use cases are related to DATA, SENSEMAKING OPERATION, SENSEMAKER and REPRESENTATATION (see Table 2). For example, SENSEMAKING OPERATIONS were expanded with operations relevant to the Dicode use cases (e.g. ABSTRACTING, CLASSIFYING, COMPARING, FILTERING,

SEARCHING, VISUALISING); and DATA were specified (e.g. STRUCTURED DATA, UNSTRUCTURED DATA, QUALITATIVE DATA, QUANTITATIVE DATA).

Table 2. Conceptualising Representation of Middle Reusable Layers

Description Logic (DL) syntax	
Representation $\sqsubseteq \exists$ typeOfRepresentation. RepresentationType	
Representation $\sqsubseteq \exists$ communicate. SharedUnderstanding	
SpatialRepresentation : RepresentationType	
FacetedRepresentation : RepresentationType	
ArgumentationalRepresentation : RepresentationType	

4.3 Lower Case Specific Layer

Three case specific ontologies were derived to capture the specificity of sensemaking activities for each Dicode use case. For example, the sensemakers in each user case are represented such as: RADIOLOGIST, RADIOGRAPHER, and CLINICAN in the RA clinical trial use case; CLINICAL RESEARCHER in the CG research use case; and MARKETING RESEARCHER in the PO use case.

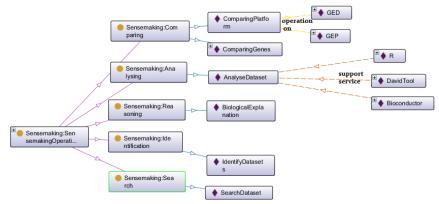


Fig. 3. Sensemaking Operations for Clinico-Genomic (CG) Use Case

Fig. 3 represents the case specific sensemaking operations related to the CG research use case (e.g. COMPARING PLAFTORM, COMPARING GENES, ANALYSE DATASET, BIOLOGICAL EXPLANATION, IDENTIFY DATASET, SEARCH DATASET) including the data such operations are performed on (e.g. GED - Gene Expression Data, GEP – Gene Expression Profile in the case of operation COMPARING PLATFORM) and support systems for such operations (e.g. R, DAVID TOOL and BIOCONDUCTOR for ANALYSING DATASET).

Concepts in the case specific layer were derived from several knowledge sources: interviews with stakeholders in each use case, relevant documentation, and user stories. The knowledge sources were analysed by a representative of domain experts following the guidance from the upper abstract layer and a knowledge glossary was built. The concepts from the glossary were then encoded in an ontology using an

intuitive ontology authoring tool ROO [19] which enables active involvement of domain experts.

Our modelling approach also allowed us to utilise relevant ontologies and datasets from Linked Data Cloud (such as DBpedia²) and public ontologies (such as RadLex³, MeSH⁴) to enhance the coverage of the concepts in DON use case modules. For example, to improve the coverage of the BODY PART concept (from the RA clinical trial use case) we utilised RadLex and MeSH ontologies (see Fig. 4).



Fig. 4. Utilising external ontologies to specialise Body Part concepts in the RA clinical trial use case (left-top: Original, left-bottom: MeSH ontology, right: RadLex ontology)

5. Utilisation of DON: Augmentor Semantic Services in Dicode & Beyond

DON ontology is being used for semantic augmentation of medical diagnosis reports and user contributions to argumentative interactions. For semantic augmentation we have developed generic services: a) Semantic Annotation service – to tag content semantically, i.e. linking content to named entities and b) Semantic Query service – to search (and to facilitate browsing of) semantically tagged content.

A web based tool *Augmentor* is developed to illustrate the utilisation of these semantic services and to understand benefits ontologies can bring. In this section, we outline the implementation details for Augmentor.

² <u>http://dbpedia.org/About</u>

³ <u>http://www.rsna.org/radlex/</u>

⁴ <u>http://www.nlm.nih.gov/mesh/</u>

5.1. Architecture and Implementation

Fig. 5 shows the main architectural elements of the currently implemented Augmentor services and their interactions.

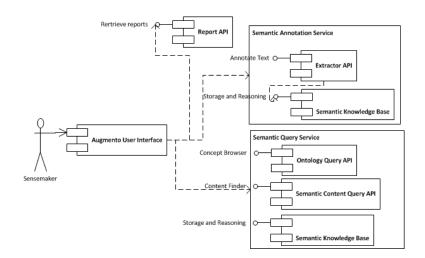


Fig. 5. UML Component Diagram for Augmentor Services

In addition to its front-end user interface, Augmentor consists of semantic annotation and semantic query services – both components are utilising DON ontologies. The interface also consumes an internal report API. The semantic reasoning and storage layer is part of the both services and works as an interface between the underlying semantic processing technologies (ontological knowledge bases, application logic, and text mining systems) and the services.

Through a URL, Augmentor retrieves metadata and selects textual content of the medical diagnosis reports kept in a web server. Semantic annotation service automatically tags content with DON concepts using text mining techniques based on GATE⁵. This service also augments the tags with the concepts from external ontologies. The content is tagged on the fly and stored in a semantic knowledge base driven by high performance OWLIM⁶ semantic repository. Browsing and retrieval of heterogeneous content (comments, metadata, and knowledge base) in semantic query service is implemented using two sets of technologies: schema level reasoner API Jena⁷ to browse the ontologies and content retrieval service based on OWLIM. The REST based implementation of these components allows utilising these services outside Augmentor user interface (see Fig. 6) by Dicode use case partners or other services in Dicode.

⁵ <u>http://gate.ac.uk/</u>

⁶ <u>http://www.ontotext.com/owlim</u>

⁷ http://jena.sourceforge.net/

In Fig. 6, four parts are being highlighted (A to D) to show the result of semantic augmentation on a medical report. *A) Comments* - the medical report contains self-reflection note/comments from a sensemaker (radiographer). These notes can be used by other sensemakers to study this sensemaker's sensemaking process while conducting clinical study. *B) Concepts* - the important concepts that describe the comments are semantically tagged and linked to the knowledge base. Clicking on the hyperlinked concepts takes the sensemakers to other related reports. *C) Sensemaking Operations* - Augmentor gives indication of the sensemakers. *D) Resources* – Augmentor interface shows the connections of the report to relevant linked datasets.

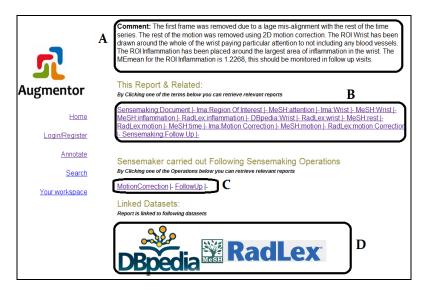


Fig. 6. Interface for Augmentor Services: a result of semantic Augmentation

5.2. Benefits of Modularisation in DON

While utilising DON in Augmentor services for semantic augmentation of content, the modularisation approach allows utilizing only relevant modules from the DON and corresponding knowledge base. It helps to constraint the annotation space for the semantic annotation service to be limited to the specific use case.

We also utilise DON in Augmentor services by providing a structure for browsing content related to sensemaking activities and knowledge bases, which can facilitate sensemakers in Dicode to take informed decisions. The browsing service requires storage and reasoning layer to support required reasoning and search functionalities. We utilise semantic technologies (e.g. semantic repositories, SPARQL⁸) to provide the storage and reasoning for developing such applications. We exploit the modularisation in DON ontologies by loading relevant ontologies and datasets to each

⁸ <u>http://www.w3.org/TR/rdf-sparql-query/</u>

use cases into separate SPARQL named graphs on semantic repositories. This allows us querying and reasoning against a subset of ontology and knowledge bases instead of the whole. Hence, we benefit from the scalability offered by the modular design.

Beyond Dicode, with reusability-driven strategy in the modularisation, we have created a set of ontologies that can be utilised and extended seamlessly in other projects and applications that focus on sensemaking activities. DON is distributed as open source⁹.

6. Conclusions & Future Work

In this paper, we have described a socio-technical approach for modelling sensemaking which is an example of cognitively complex domains. Underpinning our approach is an *a priori* modularisation methodology that enables the division of domain ontology into several modules from the outset in order to (i) systematically handle the complexity and dynamicity of ontology modelling in such domains; (ii) iteratively incorporate contributions from the social sciences into the ontology.

This approach can be followed for addressing key challenges of ontology engineering in cognitively-complex or ill-defined domains (which are becoming ubiquitous with the information explosion on the Web). In particular, utilising theoretical frameworks can be beneficial for the domain experts to guide the articulation of their understanding. We have demonstrated the application of the socio-technical approach in the context of the Dicode project where a multi-layered ontology (DON) is designed to address requirements from multiple use cases that involve sensemaking. The paper has also illustrated the use of DON in Augmentor to semantically augment and link medical diagnosis reports to assist sensemakers.

The next phase of this work includes: *a)* Further experimentations with DON in web service annotations and discovery. DON could be used as a common vocabulary between services and service developers and for enhancing the functionality of specific Dicode services such as social media mining or community mining. *b)* A user trial of DON and Augmentor. In particular, we are interested in the impact that the semantic-driven sensemaking services have in aiding sensemakers. c) Improvement to the functionalities of DON and Augmentor services. Augmentor will be further developed to cover remaining use cases from the Dicode to support clinical research professionals to make sense of scientific findings in the breast cancer domain and support market research analysts to make sense of the brand's public perceptions on social media.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no ICT 257184 (DICODE project). Thanks go to Ronald Denaux for his reviews of the paper and useful suggestions for improving the paper.

⁹ https://sites.google.com/site/ontomatic/don

References

- 1. Warren, P.: Knowledge Management and the Semantic Web: From Scenario to Technology. IEEE Intelligent Systems, IEEE Intelligent Systems, 21, pp. 53-59 (2006)
- Stojanovic, N. and Handschuhs, S.: A Framework for Knowledge Management on the Semantic web, In: 11th International World Wide Web Conference, Honolulu, Hawaii, USA (2002)
- 3. Watanabe, K.: Introduction of Dublin Core metadata. Journal of Information Processing and Management, 43 (2001)
- Harris, M.A., Clark, J.I., Ireland, A Lomax, J. and Ashburner, J.: The Gene Ontology (GO) project in 2006. Nucleic Acids Research, 34, pp. 322-326 (2006)
- Rector, A. and Rogers, J.: Ontological Issues in using a Description Logic to Represent Medical Concepts: Experience from GALEN. IMIA WG6 Workshop: Terminology and Natural Language in Medicine. Phoenix Arixona (1999)
- Russell, D. M., Stefik, M. J., Pirolli, P. and Card, S. K.: The cost structure of sensemaking. In INTERCHI '93 Conference on Human Factors in Computing Systems, Amsterdam (1993)
- Dervin, B.: Sense-Making Theory and Practice: An overview of user interests in Knowledge seeking and use. Journal of Knowledge Management, 2(2), pp. 36-46 (1998)
- Pirolli, P. and Card, S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proceedings of International Conference on Intelligence Analysis (2005)
- Whittaker, S.: Making sense of sensemaking. In T. Erickson and D.W. McDonald (Eds.): HCI remixed: Reflections on works that have influenced the HCI community, pp. 173–178 (2008)
- 10.Omitola, T., Millard, I., Glaser, H., Gibbins, N. and Shadbolt, N: From Information to Sense-Making: Fetching and Querying Semantic Repositories. In: KES 2010, Part IV, Lecture Notes in Artificial Intelligence, 6279, Springer (2010)
- 11.Dadzie A-S, Iria, J., Petrelli, D. and Xia L: The xmediabox: Sensemaking through the use of knowledge lenses. In Extended Semantic Web Conference, pp. 811-815 (2009)
- Ure, J., Proctor, R. Data Integration in eHealth: A Domain/Disease Specific Roadmap, Studies Health Technolgy Information, pp. 144-53 (2007)
- 13.Trist, E and Bamforth, K.: Some social and psychological consequences of the longwall method of coal getting, in: Human Relations, 4 (1), pp. 3-38 (1951)
- 14.Clegg, C.W.: Sociotechnical Principles for Systems Design, Applied Ergonomics, 31, pp. 463-477, (2000)
- 15. Scacchi, W.: Socio-technical design, In: Bainbrigde, W.S. (Ed.), The Encyclopaedia of Human-Computer Interaction, Berkshire Publishing Group (2004)
- 16.Thakker, D., Dimitrova, V., Lau, L., Denaux, R., Karanasios, S. and Yang-Turner, F: A Priori Ontology Modularisation in Ill-defined Domains. Accepted for the I-Semantics 2011: 7th International Conference on Semantic Systems. Graz, Austria (2011)
- 17.Simperl, E.: Reusing ontologies on the Semantic Web: A feasibility study. Data Knowledge Engineering. 68(10), pp. 905-925 (2009)
- 18.Paul, S.A. and Reddy, M.: A Framework for Sensemaking in Collaborative Information Seeking. Proceedings of 2nd International Workshop on Collaborative Information Seeking at CSCW 2010, Savannah, GA (2010)
- 19.Denaux, R., Dolbear, C. Hart, G., Dimitrova, V. and Cohn, A.G.: Supporting Domain Experts to Construct Conceptual Ontologies: A Holistic Approach. Web Semantics Science Services and Agents on the World Wide Web, 9(2), pp. 113-127 (2011)
- 20.Pirolli, P. and Russell, D. M.: Introduction to this special issue on sensemaking, Human-Computer Interaction Journal, 26(1-2), pp. 1-8 (2011)