

From the knowledge acquisition bottleneck to the knowledge acquisition overflow: A brief French history of knowledge acquisition

Nathalie Aussenac-Gilles, Fabien Gandon

▶ **To cite this version:**

Nathalie Aussenac-Gilles, Fabien Gandon. From the knowledge acquisition bottleneck to the knowledge acquisition overflow: A brief French history of knowledge acquisition. International Journal of Human-Computer Studies, 2013, Special issue: 25 Years of Knowledge Acquisition, 71 (2), pp.157-165. 10.1016/j.ijhcs.2012.10.009 . hal-01124416

HAL Id: hal-01124416
https://hal.science/hal-01124416

Submitted on 6 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 12333

To link to this article : DOI :10.1016/j.ijhcs.2012.10.009
URL : http://dx.doi.org/10.1016/j.ijhcs.2012.10.009

To cite this version : Aussenac-Gilles, Nathalie and Gandon, Fabien *From the knowledge acquisition bottleneck to the knowledge acquisition overflow: A brief French history of knowledge acquisition*. (2013) International Journal of Human-Computer Studies, vol. 71 (n° 2). pp. 157-165. ISSN 1071-5819

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

From the knowledge acquisition bottleneck to the knowledge acquisition overflow: A brief French history of knowledge acquisition

Nathalie Aussenac-Gilles[a,*], Fabien Gandon[b]

[a]*IRIT - CNRS, Toulouse, France*
[b]*INRIA Wimmics, Sophia Antipolis, France*

Abstract

This article is an account of the evolution of the French-speaking research community on knowledge acquisition and knowledge modelling echoing the complex and cross-disciplinary trajectory of the field. In particular, it reports the most significant steps in the parallel evolution of the web and the knowledge acquisition paradigm, which finally converged with the project of a semantic web. As a consequence of the huge amount of available data in the web, a paradigm shift occurred in the domain, from knowledge-intensive problem solving to large-scale data acquisition and management. We also pay a tribute to Rose Dieng, one of the pioneers of this research community.

Keywords: Knowledge modelling; French research; Knowledge-based systems

In this article we give a localized account of the evolution of the domain of knowledge acquisition (KA) that Brian Gaines has presented in a broader perspective in his contribution to this special issue (Gaines, this issue). We contrast the evolution of KA with the parallel evolution of the Web and indeed, in the last 10 years, knowledge engineering as a research domain and the Web have converged in particular in the Semantic Web project and the current Web of Data. Here we describe the evolution of the French-speaking conference on knowledge acquisition and knowledge modelling in order to give an overview and a brief history of the domain, which has at times been hidden by the language barrier. This evolution echoes the complex and cross-disciplinary trajectory of the field presented in Brian Gaines' exhaustive outline and is also consistent with Musen's historical outlook on the last 25 years of the international workshops (Musen, this issue). At times one or two years earlier or later, the French KA conferences reveal a strong convergence and consistency with the research trends and paradigm shifts which have characterized the KA domain.

With this paper, we also wish to pay tribute to the generosity, the scientific talent and the unforgettable smile of one of the pioneers in the French and international KA research communities, our colleague Rose Dieng.

1. When AI requires knowledge acquisition

The French AI scientific groups interested in building rule-based systems or learning systems highlighted knowledge acquisition as a research issue as early as 1986. Pioneer researchers and engineers from innovative companies experimenting expert systems (like CEA and EDF) organized an informal scientific meeting in 1988, whereas J.G. Ganascia and Y. Kodratoff co-chaired one of the first EKAW workshops in Paris in 1989.

The French-speaking conference on knowledge engineering is called IC, for *Ingénierie des Connaissances* (*Knowledge Engineering*). It started soon after, in 1990 and was known at the time as the JAC, for *Journée d'Acquisition des Connaissances*, literally the knowledge acquisition day. Approximately at the same time the idea of Web was born

*Corresponding author.
E-mail addresses: aussenac@irit.fr (N. Aussenac-Gilles), Fabien.Gandon@inria.fr (F. Gandon).

at CERN. The acquisition problem is directly inherited from expert systems and focuses on capturing the knowledge needed to feed them. Research on Knowledge Acquisition was one of the two trends motivated by the limitations faced by expert systems, the other one being the definition of richer logic-based knowledge representations.

In 1990, the first JAC (Knowledge Acquisition Day) was organized by CNET in Lannion (Centre National d'Etudes en Télécommunications) following a meeting on the same topic organized by the Artificial Intelligence research group GDR-PRC in January 4, 1989. At the time, the JAC aimed at gathering the francophone community in the field of knowledge acquisition and to clarify the relationship between this domain and machine learning. This first edition was on purpose positioned within a multidisciplinary framework, as shown by the diverse origins of the presentations that day, including computer science, industrial research, and psychology. The French knowledge engineering community (*IC*) still maintains as a birth mark this specificity to be a multidisciplinary conference, rather than a sub-domain of Artificial Intelligence. Over the years, the disciplines involved in knowledge engineering have changed in keeping with the research main trends. In 1990, the strong reference to structuralism in expert systems assumed that rules and frames are more than convenient implementation paradigms: they have a cognitive validity and reflect cognitive structures. Expert systems map human expertise, often the one of a single expert, and they aim to solve problems using the same heuristics as the expert.

The research issues formulated at the time determined most of the domain structuring paradigms for the next ten years: What is the *right abstraction level* to describe the system problem solving behaviour independently of the formal representation? *How can a conceptual model guide the identification of the knowledge to be captured* in the system? What is the structure and content of these models? And finally which formalisms should be used and which processes should be supported? Acquisition methods included interviews as well as psychological techniques like card sorting or the repertory grids inspired from Kelly's Personal Construct Theory and promoted by Boose (1984), Gaines and Shaw (1980). In France, AI researchers collaborated with cognitive psychologists to define acquisition methods, one of the most famous being KOD by Vogel (1988).

The same year, Tim Berners-Lee proposed for the second time at CERN a memo specifying a system he first called "Mesh", where he suggested the use of hypertext for information management within CERN, by extending the references of hyperlinks to network addresses of documents in order to build a "mesh" between documents stored on different machines. The memo was returned to him with the handwritten "vague but exciting" and so began the development of the technical architecture of the Web. This technology will in itself revolutionize the problem of knowledge acquisition and knowledge publishing by extending Ted Nelson's hypertext principle (Nelson, 1965) of a "file structure for the complex, the changing, and the indeterminate" information to the scale of the Internet, creating a whole new social medium of knowledge.

In 1991 *la JAC* became *les JAC* (Knowledge Acquisition Days) spanning several days and the program focused on the methodologies for the *acquisition and modelling* of knowledge including references to methods like KADS (Born, 1990) and KOD (Vogel, 1988). Newell's knowledge level (Newell, 1980) was systematically cited as the right one to describe problem solving knowledge before its encoding into a concrete symbol-level representation. The first established results of a large European project named KADS suggested that several description layers were required to analyse the system knowledge from various perspectives. Influenced by Chandrasekaran's (1986) generic Tasks and McDermott group's work on Role Limiting Methods (McDermott, 1988) (reusable problem solving models), KADS clarified how the system goals and tasks differed from the processes and methods followed to carry out the task. So, one of the most studied layers was the problem solving model. It described reasoning methods independently of the domain knowledge, only taking the type of problem into account. The two foundational research issues here are *knowledge reuse* (how much can a knowledge model be reused and how? what is reusable in a knowledge model?) as well as the *definition of modelling primitives* (what are the components of a conceptual model at each layer? how are these components and layers linked together?).

That same year, the first Web server was installed outside Europe.

2. Knowledge acquisition for modelling, modelling to guide acquisition

In 1992 the JAC (Knowledge Acquisition Days) were particularly interested in the analysis of textual corpora, through natural language processing, to acquire knowledge. Another key issue was the nature and reuse of knowledge components for generic models. The emergence of issues related to the exploration of textual documents as knowledge source resulted from the systematic rewriting of experts' interviews. Language analysis was basic and reflected the simplistic semantic hypotheses formulated at that time: each sentence was expected to provide various predicates and logic formulae.

The then famous KADS's four-layers modelled to several significant changes in the role of the conceptual model: firstly, a conceptual model no longer modelled an expert knowledge, it represented the system knowledge; secondly, the model could be used as a grid that drove the acquisition process. Once selected and adapted, a problem solving method determined the roles played by domain knowledge and identified various sets of rules to be applied at each step of the method. Model-based KA opened new perspectives to building knowledge-based systems. Whereas the "Knowledge Acquisition Journal" first issue contributed to state that KA was not just a short-term problem, but rather a complex

research issue that would survive rule-based systems, a special issue of the RIA[1] French journal (Aussenac-Gilles et al., 1992) made this research domain more visible in the French AI landscape. In 1988, a French special interest group about KA (called GRACQ[2]) was set up with the help of the CNRS and the French association of AI (AFIA). It gathered most of the research teams working in this area. GRACQ organized 6 annual meetings for several subgroups dedicated to problem solving methods, control knowledge, reuse, text analysis, extraction techniques and methodological issues.

There were then a dozen web servers and new Web browsers appeared in the course of the year.

The (so often cited) paper (Gruber, 1991) by T. Gruber about ontology issued in 1991 suggested that an ontology can be used as a model to represent domain knowledge in a reusable way for various applications. Gruber's paper managed to overcome a double difficulty: to turn a complex philosophical notion into a rather easy to understand technical object; and to solve two reusability issues (format standardization and content unification, consensus within a domain) with one single model. It would take 2 years before this paper impacted on the French knowledge acquisition community.

In 1993, JAC was not organized because the European equivalent event, the EKAW'93 conference, took place in France for the second time. The numerous activities of GRACQ were presented via posters during the open day, demonstrating the dynamism of more than 30 research groups.

The domain continued to evolve in keeping with new perspectives on knowledge-based systems (KBS): knowledge sources were diversified, the experts' know-how was better characterized and collaborations with human factor specialists underlined the necessity to capture knowledge in use, to focus on the actual human activity and not only on their discourse about it. Rather than independent problem solvers, knowledge-based systems were defined as users' assistants that carry out only a part of the overall task. Feedbacks from various attempts to provide users with expert systems proved that the quality of the expert system solving capabilities does not guarantee that the system is useful for its users. Human-centered design (Norman, 1986) suggested that the focus was not the system on its own but the pair (user, system) and its interactions. Zacklad (1993)[3] by is one of the first papers at JAC to explore what a knowledge-based interactive support system could be.

That same year, CERN leaders officially announced that Web technology would be free and royalty free. This important step would allow viral penetration of these technologies in other information systems. Earlier this year, there were about fifty servers. New browsers appeared but the most important one was Mosaic which allowed for the first time to visualize the images directly in the text of a page. With this browser, the Web would actually spread around the world leaving behind his ancestors Gopher, WAIS, and FTP. The so-called CGI (Common Gateway Interface) was also proposed to allow web servers to not simply send static pages but also to execute a program to return a generated content. This technical detail would open a huge new avenue for the Web, enabling it to go beyond the documentary service to provide the means of universal access to service applications.

In 1994, the program of the JAC included a new word alongside knowledge models: ''ontology''. A single paper examined the notion of ontology (Reynaud and Tort, 1994)[4], and explained to the KA French researchers what makes ontologies different from domain models. Hot topics were modelling methods, model validation and simulation, the use of cases in knowledge models, the comparison of methods and the feedback analysis of their experimental use. Another original contribution (Lepine and Aussenac-Gilles, 1994) was dedicated to an innovative way to exploit texts as knowledge sources: instead of a systematic sentence to model mapping, the idea was to use a term extractor and build up a domain vocabulary to speed up the modelling process.

The Web browser Mosaic was spreading fast and researchers in the French community wondered what would be the impact of this tool compared to the Minitel[5] system which had been deployed in France. More than 600 Web servers were now online, but more importantly, the first edition of the World Wide Web conference (WWW 1994[6]) invited Tim Berners-Lee to present his vision of what would become the Semantic Web. In parallel, with the proliferation of browsers the ''browser war'' began and to avoid the dangers of fragmentation or monopoly, a standardization body was created for the Web: the W3C.

3. From knowledge in use to models: knowledge engineering

In 1995, alongside topics now classic at JAC like knowledge acquisition, explanation, and modelling and representation languages, two new topics first appeared, which would remain over the following years: (1) *return on experience*, including the problem of integration with legacy systems and (2) bidirectional links between knowledge *modelling* on the one hand and *learning* and reasoning on the other hand. 1995 was the year when the CommonKADS (Schreiber et al., 1999) project started as a revision of KADS that would take into account the context in which a system is used, task sharing between the user and the KBS, and the possible conflicts and divergences between knowledge sources. CommonKADS also aimed

[1] *Revue d'Intelligence Artificielle* (AI Journal).

[2] GRACQ stands for Groupe de recherche ACQuisition des Connaissances. Founders and early leaders of this group are J.P. Krivine, P. Laublet, N. Aussenac-Gilles, G. Kassel, J. Charlet and R. Dieng. http://www.irit.fr/GRACQ/

[3] Read (Zacklad and Rousseaux, 1996) for an English presentation of this work

[4] early ideas that were later developed in (Reynaud and Tort, 1996)

[5] http://en.wikipedia.org/wiki/Minitel

[6] http://www94.web.cern.ch/www94/

at more flexibility in the definition of specific problem solving methods from generic and reusable ones.

More than 10,000 Web servers were then available. By introducing CSS, the W3C began looking at how to standardize the decoupling of Web content and its presentation in particular to facilitate multiple processing of the same content. By separating the content structure from the presentation format the web made a first step toward a web of structured content.

In 1996, the *JAC* program highlights were: ontologies, refinement of models and methods, cooperative system construction and industrialization of approaches. One of the challenges was to push KA out of the academic area and to be able to demonstrate the feasibility of these approaches when dealing with "real" applications. Worrying about actual applications and not only toy data-sets was one of the key features that implied changing from *Knowledge Acquisition* to *Knowledge Engineering* (KE). The complexity of knowledge and contextual features to be taken into account in industrial applications required combining in an innovative way more basic or elementary contributions, techniques and tools. The technical feasibility of this combination raised engineering issues, and its validity, given the theoretical background and methodological assumptions behind each technique, was a true research problem per se, at the frontier of Knowledge Engineering.

There were at the time more than 100,000 Web servers. This year also saw one of the first actions on web security with PICS recommendation for the protection of children against inappropriate content: the size and the diversity of the content of the web were starting to create new challenges in parallel with their growth.

In 1997, JICAA (Days of Knowledge Engineering and Machine Learning) is the hinge between JAC (Knowledge Acquisition Days) and IC (Knowledge Engineering conference) showing how the challenges are broadening from the problem of acquisition. KBS were by then part of the broader context of information systems, and knowledge acquisition and validation had to be positioned within knowledge engineering. The trend was to open conferences to very different areas so that theoretical and applied concerns could coexist with a maximum of relevant different approaches. Among the session topics, we find ontology, terminology and knowledge extraction from text. The switch observed in 1994 about text analysis became obvious. The use of mining techniques to explore text and extract linguistics clues from which pieces of model could be built up defined a new trend. It opened the way to designing, combining and testing innovative techniques for Natural Language Processing and knowledge modelling. Additional topics were: hypertext and document modelling, corporate memories and collaborative systems, validation of knowledge for KBS, data mining, case-based reasoning, and cognitive approaches. As an echo to Clancey's notion of "situated cognition" (Clancey, 1997), the challenge addressed here was the integration of the social and collective dimension of knowledge in the definition of an information system. The information system not only included a KBS, but it also defined a work organisation that distributed knowledge and problem solving over the system, its users, documents, and other human activities. A debate arose concerning the scope of KE: Should it systematically include a technical and engineering AI-based contribution? Should it systematically require the integration of human factors and organizational issues? This debate is an on-going one, although the strong influence of computer science in the field often sets the technical issues at first place.

4. Knowledge is distributed, collective and embedded in activity

In 1998, the word 'Web' appeared for the first time in the title of an article at IC (Knowledge Engineering conference). It was about time since the Web had already passed the one million servers mark the same year. While knowledge representation, modelling languages and corporate memories were still very present in IC'98 program, hypertext and hypermedia also confirmed their importance as a means to support access to knowledge in information systems. It was admitted that knowledge-based systems do not necessarily solve problems or use logic languages to produce inferences. Knowledge systems provide a more or less dynamic access to knowledge as it can be captured by language in text, by use traces or cases. A session was also dedicated to computer mediated cooperative information systems. Vincent Quint gave an invited talk about going beyond HTML to rich documents on the Web. Indeed the same year the first recommendation on XML was published. The other invited keynote from Richard Benjamins opened up perspectives on how to transform part of the Web into a knowledge base. That same year, Tim Berners-Lee published a paper called "Semantic Web Road map" (Berners-Lee, 1998).

In the KE domain, CommonKADS was about to end. Many papers referred to its use, evaluation or adaptation to specific domain data. Modelling the problem solving knowledge was the core issue that had been significantly improved with regard to KADS. Rather than a library of static methods, CommonKADS offered a library of adaptable components and, for each of them a set of questions. Answering these questions guided the selection of the best adapted components that would form the system problem solving model. CommonKADS also provided modelling languages to represent a variety of forms of knowledge including problem solving methods, domain knowledge and ontologies, and traces of activity tasks.

More generally, KE research issues concerned both technological and representation issues, as well as organisation and human factors. Methodological questions like how to access collaborative knowledge, knowledge in use and practices in organizations needed to call for concepts, techniques or theories (like complexity theory and systemic

approaches) from other disciplines, such as, management, sociology or ergonomics. The scope of knowledge engineering was widening and contributions explored very different and complementary questions in the process that ranges from qualifying source knowledge in human activities to making it operational or available via an application.

In 1999, IC (Knowledge Engineering conference) still had sessions on ontologies, and knowledge acquisition from text but a new theme was introduced on cognitive ergonomics and knowledge engineering vs. requirement engineering. A particular problem raised was the one of the links between the study of information systems, the engineering of knowledge-based systems and the management of organizations in which these systems are deployed. The challenge was to integrate the fact that *our systems are immersed in usage*. The conference papers showed that the introduction of new processes leveraging individual and collective knowledge raised a set of interrelated problems that could no longer be treated separately but had to be addressed jointly by computer scientists, cognitive scientists, researchers in ergonomics, sociologists and managers.

W3C working groups at this time generalized the integration of multimedia objects and scripts in Web pages, which would allow the Web to acquire richer, more dynamic and more reactive content. These were the basic techniques that enabled the birth of Web 2.0 in the following years. In turn this new richness of multimedia content and interaction means would tremendously change and augment the content of the Web and the applications it would allow.

5. From problem solving systems to knowledge browsing and model querying

In 2000, the emphasis in IC was on Intranet and Internet, information systems and education engineering. An important part was also given to proofs of concepts, through a demonstration session with 20 software applications and two parallel sessions on important achievements in the field. A session was dedicated to the epistemology of knowledge engineering and another session was concerned with the links between hypertext and ontologies. Terminology and knowledge extraction from texts were still core topics of the program but we also discussed the evolution of the ''knowledge'' capital in a business, object-oriented modelling of knowledge and knowledge management. The role of conceptual models had then definitively changed: they were no longer just intermediary representations before operationalisation, but rather representations to be browsed, queried or used to support man-system cooperation and task distribution. In this context, graphs emerged as a very rich and relevant knowledge representation. In particular, conceptual graphs, defined at first to represent natural language interpretations, were used to operationalize the first versions of the Semantic Web languages defined by the W3C. RDF and RDFs standards proposed to structure knowledge in triples that connect URIs and values with labelled relations. The set of triples in a domain share some nodes and form a complex graph. Reasoning with graphs, querying graphs of instances with engines like Corese (Corby et al., 2004) were one of the important research issues that emerged then. That same year, Rose Dieng published a book in French with her team on methods and tools for knowledge management (Dieng-Kuntz et al., 2000) while chairing the International Conference EKAW 2000.

In 2001, the Semantic Web appeared in the program of IC. It would be the theme of a session in addition to other topics including: Information Systems for Decision Support, epistemology, practical experiences, cooperation and case-based reasoning.

The collective dimension of knowledge requires one to define how to combine various knowledge sources when building an ontology. For instance, the ontology building platform Protégé was modified to support the collaborative design of ontologies (Tudorache et al., 2008), thus allowing different kinds of organizations that distribute roles and rights to contribute to the ontology, and to check and validate these contributions. Another emerging challenge was to optimize knowledge reuse and model combination.

The Web was unavoidable with 26 million servers. W3C groups were interested in generalizing the notion of link in structured documents (XLink) and in allowing access to the Web by voice and hearing (Voice XML), including the phone. These changes prefigured the ideas of linked data and mobile web.

In 2002, IC was still marked by multidisciplinary contributions and topics as diverse as: problem solving systems, hypermedia systems, education engineering, terminology extraction from texts and knowledge management. Several contributions compared these paradigms and stressed how closely the issues raised by each paradigm were related to knowledge management. Companion disciplines were natural language processing and computational linguistics, human factors and distance learning. The topic of ontologies had become central. The Semantic Web was also now a topic of its own. Industrial design emerged as a new topic in the application sessions. An intelligent system was no longer an active problem solver but rather an assistant that drives the user towards the relevant information that will assist him properly. Navigation through information sources thanks to semantic annotations turned out to be one of the most popular knowledge-based applications. After the pioneering work by Guarino and Masolo (Guarino et al., 1999) ontologies were used for concept-based annotation of domain specific documents. The larger the document collections, the more generic the ontology. The popularity of lexical databases like WordNet, and their limitations (Gangemi et al., 2003), suggested the need for developing more structured and appropriate ontologies.

At the same time, W3C offered a first recommendation (P3P) to promote respect for the privacy of users and these aspects were starting to worry more and more people in the

community. In parallel, another group working on web accessibility provided good practices to ensure that knowledge on the web remained accessible.

In 2003, among the topics we could find modelling, text and ontologies, Semantic Web, knowledge management, design and implementation. Semantic web applications were mature enough to run experiments where the gain brought to users could be qualitatively (if not quantitatively) qualified. The social dimension of the annotation processes and ontology definition was one of the few references to human factors at this conference. The increasing number of technical issues, related to ontology mapping, concept-language association and the improvement of reasoning facilities, became a major focus that put aside the study of human acceptance and use of these technologies. The W3C proposed an evolution of Web forms that could be generalized to the whole family of XML languages. The web continued to host more and more applications and not only content.

6. A world of Ontologies

In 2004, IC considered the relationship between knowledge engineering and document engineering, cooperative activities and support systems. In addition to text mining, the topic of data mining was added, whereas it was addressed mainly in other conferences up to that time. The popularity of mining approaches to acquire knowledge models revealed the increasing efficiency of machine learning and statistics to ''learn'' the rules and knowledge revealed by data regularity. Mining and learning took advantage of the increasing number of available electronic textual documents. These methods competed with (and often out-performed) the manual analysis of human expertise and expert documents. There was also a session on conflict resolution and consensus building, which confirmed that knowledge was diverse and that unifying models had to cope with this diversity. A panel reported on the transfer to industry of knowledge technologies. Jérôme Euzenat's invited talk introduced OWL, the official recommendation from W3C to extend the expressiveness of the Semantic Web formalisms.

The number of Web servers exceeded 46 million. Web access from mobile phones and PDA had now become an important activity in W3C.

In 2005, ontologies attracted more and more research work, either from a technical point of view or regarding their use and engineering. The IC conference managed to bring together the issues of building and using ontologies on the one hand and of knowledge engineering in organizations on the other hand. Several papers provided methodological proposals for the construction of ontologies from text corpus or from existing knowledge bases. There was a development of the topic of the alignment of existing ontologies. Indexing and annotation using ontologies for intelligent information search were also well presented in this edition. Several articles illustrated the openness to other disciplines such as organization theory, cooperative work systems and education engineering.

Web addresses (URLs) became multilingual (IRI), as characters other than ASCII ones could now be used in addresses. W3C also launched an initiative to promote the Web in developing countries and a working group was created to work on the contributions of the Semantic Web to the medical field.

In 2006, annotations were an important topic of IC, as well as mapping and visualization of knowledge. These research issues were directly linked to the increasing amount of information and documents available on the web. Existing wikis, tagged documents and tag hierarchies revealed new types of ''social'' knowledge. They provided valuable resources to identify usage-driven relationships between concepts. As the web grew in size and diversity, the challenge to turn it into a semantic web became more complex and included the management of large data-sets, as well as the access to text and document content. Concept-based annotation combined the use of information extraction and NLP techniques. Once annotations are available, new browsing and reading devices can be designed that take benefit from additional knowledge expressed in either natural language or a formal representation. The question ''Can the Semantic Web be social?'' was raised for debate (Zacklad, 1999; Gandon, 2006): beyond the technical issues, beyond the seduction of the semantic web infrastructure, how can ontologies and ontology-based annotations take into account the diversity of social groups and their points of view? What are the social and technical challenges to be addressed before this technology gets a real take-up from users?

The issue of internationalization spanned several working groups at W3C, which also inaugurated an office in China. Internationalization addressed cultural differences, language and domain specificities.

In 2007, IC discussed the Semantic Web, its complementarities with the Web 2.0 and concept-based search for information, with a focus on the synergy between annotations and knowledge-based systems. Other topics included the analysis of texts and ontologies, applications of knowledge engineering, as well as cooperation and sharing of knowledge within human communities. During this same year, the FreeBase[7] database, described as ''an open shared database of the world's knowledge'' was made available. Tim O'Reilly defined it as ''the bridge between the bottom up vision of Web 2.0 collective intelligence and the more structured world of the semantic web.'' Another major community project was launched: the first version of the DBpedia[8] ontology was built from Wikipedia info-boxes (Auer and Lehmann, 2007). The same year appeared the first Linked Open Data cloud on the web, which was based on metadata collected and curated by contributors to the CKAN directory. This cloud would mirror in the following

[7] http://wiki.freebase.com/wiki/Main_Page
[8] http://dbpedia.org

years the growth of available linked data on the web, each dataset acquired from very different sources.

W3C, finally adopted a query language for XML (XQuery) and proposed (GRDDL) as a gateway between the structured Web (XML documents) and the Semantic Web (RDF graphs). W3C opened an office in South Africa reminding us that the digital divide was real and that the progresses we have made on knowledge engineering and the semantic web sometimes could lead to the exclusion of even more users, who could not benefit from them.

7. The semantic web: turning data into knowledge

In 2008, IC sessions included the dissemination of medical knowledge, querying knowledge graphs, ontological knowledge-based systems and ontology design. The invited speaker Ivan Herman, Head of the W3C Semantic Web Activity, gave an overview of the Semantic Web that led to a very rich discussion on the relationships between the Semantic Web and the Social Web. A session on Semantic Web and Web 2.0 echoed a co-located workshop "IC 2.0", considering advances towards a social knowledge engineering process, which used Web 2.0 to change knowledge engineering practices. Other sessions addressed the extraction of knowledge from texts, knowledge-based instrumentation of practices, traces and inscriptions of knowledge and learning and adaptation. Another debate concerned the question of the value of a French-speaking Knowledge Engineering conference in its relations with the international conferences in the same domain.

On the Web, a very large number of ontologies was then available, which could be accessed thanks to semantic engines like Watson[9], Swoogle[10] or Sindice[11]. The majority of these ontologies contained concepts with labels in English or no label at all. When trying to semantically index text documents in languages other than English, several difficulties arose to select a relevant ontology: the lack of resources in the same language as the document, the complexity of understanding the cultural background of an ontology and the need to check whether the points of view are the same in the ontology and the documents. Research on ontology localization attempted to define rich representations for multilingual or cross-cultural ontologies. The question here covers a general issue that has to be dealt with if knowledge models are to be widely spread: How can one ensure the correct interpretation and local adaptation of an ontology? What are similar ontologies? How can we appreciate the adequacy of an ontology to index a document?

In 2009, IC took place in Tunisia, outside France for the first time. The conference had a special session on the topic "Knowledge and online communities" and we can find in the programme issues of content, construction, life cycle and population of ontologies. Ontology-based annotation was again present as well as the design of interfaces and interactions. In addition, two original sessions were added: a session dedicated to the evaluation of semantic similarity and the adaptation of ontologies to the user and a session devoted to modelling processes, practices and cases. The notion of ontology design pattern, defined in the NeOn[12] European project, adapted the software engineering notion of design pattern to ontology engineering so that ontology structures could become more reliable and reflect the actual meaning that ontologist has originally assigned to them.

In 2010, IC promoted emerging issues thanks to four workshops (such as, medical semantic web and ontology evaluation) and a tutorial on modelling before formalizing ontologies. Rather than technical, current issues were methodological ones. Many papers focused on how to evaluate the quality and relevance of these data, and how to adapt models and metadata to specific applications and users. A striking evolution was the convergence between knowledge engineering and information extraction and information retrieval. Although evaluation criteria remained different in these scientific fields, the perspective of the semantic web offered a ground for collaborations. Concept-based and ontology-based annotations were experimented as an alternative to current search engines. The expected gains are not that easy to demonstrate. Evaluation would require to improve word matching, to be able to evaluate synonymy and to define measures that compare concepts from one or several ontologies. All these research issues define some of the recurrent topics of the last four years. Another consequence of the large amount of web documents was the increasing number of references to learning, either through Natural Language Processing or through activity analysis on the basis of digital traces.

In 2011, presentations at IC confirmed the importance of building ontological reference models and their complementarities, as knowledge sources, with human expertise, text, and social organisation. New issues must be dealt with as large ontologies are now used in dynamic contexts: ontology modularity and ontology evolution. An important application domain that emerged recently is the management of business rules, their maintenance, validation and use to generate documents. Domain ontologies integrate formal representations of such rules as constraints on domain concepts. The web has by now gathered a large variety of knowledge structures, of different nature and form. More than a third of the papers explored the use of ontologies for applications as diverse as management of geographic knowledge, map legend design and temporal annotations.

8. Research challenges for semantic applications

This year at **IC 2012** web 2.0 and social web had a whole session, while the semantic web had two sessions in

[9] http://kmi-web05.open.ac.uk/WatsonWUI/

[10] http://swoogle.umbc.edu/

[11] http://sindice.com/

[12] http://www.neon-project.org/

addition to ontology design, semantic annotation, ontologies, e-learning, terminology, and data mining. Invited speakers addressed two topics: conceptualizations in the domain of law, rights and legal matters; and the analysis of complex networks to explore interaction networks using their community structure. The questions related to ontology building and use had a strong continuity with the early KA issues: Which kind of knowledge can be modelled? How important is the distortion between the computer interpretation of an ontology and the human interpretation of the associated knowledge? What kind of intelligent behaviour and tasks can formal representations be useful for?

Clearly, the shift from early works comes from the amount of data available as sources: the current hypothesis is that quantity and redundancy of basic formal information (like RDFs triples in the web of data) will balance the lack of precision and complexity. The initial problem of KA of having too few sources has shifted towards the problem of too many sources, too big, too diverse, too dynamic, too distributed, etc. New challenges arise in addressing data and schema heterogeneity and provenance, incoherence, and social life cycles. Old challenges are increased tenfold by the scale and complexity of the web (acquisition, formalisation, evaluation, evolution, etc.).

Moreover, as the web becomes ubiquitous, not only do we deploy a ubiquitous means of access to knowledge, we also deploy a ubiquitous means of acquisition through a mobile web, a web of sensors, a web of things. And the long tail of knowledge resources on the web requires many of the advances established by knowledge engineering, in order to be fully exploited, if we want knowledge and intelligence to emerge from the overall graphs they weave. In accordance with Wielinga (this issue), we can conclude that the results and experience acquired in knowledge engineering in the past 20 years will be important to avoid repeating with the web of data the same errors that were made when trying to model human expertise. The current standardization efforts provide a unique syntactical access to data but they do not necessarily make them ''understandable'' or semantically manageable by a software program. In other words, the deployment of the web of data is only the first phase of the deployment of a fully-fledged semantic web, a fully-fledged web of acquired knowledge.

References

Aussenac-Gilles, N., Krivine, J.P., Sallantin, J., 1992. Numéro spécial acquisition des connaissances. Revue d'Intelligence Artificielle (RIA), Hermès, Paris 6 (2), 1992.

Auer S., Lehmann J. 2007. What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content. In Franconi et al. (eds), *Proceedings of European Semantic Web Conference* (*ESWC'07*), LNCS 4519, pp. 503–517, Springer.

Berners-Lee T. 1998, Semantic Web Road map, W3C Note, 〈http://www.w3.org/DesignIssues/Semantic.html〉.

Boose J., 1984, Personal construct theory and the transfer of human expertise. In: Proceedings of the 6th European Conference on Artificial Intelligence ECAI'84, Advances in Artificial Intelligence. T. O'Shea Ed.: 51–60.

Born, G. 1990. KADS: a methodology for developing large AI systems, *CompEuro '90.* In: *Proceedings of the 1990 IEEE International Conference on Computer Systems and Software Engineering*, pp.166–171, http://dx.doi.org/10.1109/CMPEUR.1990.113622.

Chandrasekaran, B., 1986. Generic tasks in knowledge-based reasoning: high-level building blocks for expert system design. IEEE Expert Autumn 86, 23–30.

Clancey, W.J., 1997. Situated Cognition: on Human Knowledge and Computer Representations. Cambridge University Press.

Corby O., Dieng R. and Faron-Zucker C., 2004 Querying the Semantic Web with Corese Search Engine, In: Proceedings of Prestigious Applications of Intelligent Systems PAIS, ECAI, Valencia.

Dieng-Kuntz, R., Corby, O., Gandon, F., Giboin, A., Golebiowska, J., Matta, N., Ribière, M., 2000. Knowledge Management: Méthodes et Outils Pour la Gestion des Connaissances, InfoPro. Management des systèmes d'information, Dunod.

Gaines, B.R., Shaw, M.L.G., 1980. New directions in the analysis and interactive elicitation of personal construct systems. International Journal Man-Machine Studies 13, 81–116.

Gaines, B.R. Knowledge acquisition: past, present, and future. International Journal of Human–Computer Studies, http://dx.doi.org/10.1016/j.ijhcs.2012.10.010, this issue.

Gandon F., 2006. Le Web sémantique n'est pas antisocial, In: *Proceedings Ingénierie des Connaissances, Semaine de la Connaissance, Nantes*, pp. 131–140, 20–30 Juin 2006.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., 2003. Sweetening WORDNET with DOLCE. AI Magazine 24 (3), 13–24.

Gruber, T. R. 1991. The role of common ontology in achieving sharable, reusable knowledge bases. In J.A. Allen, R. Fikes, & E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Cambridge, MA*, pp. 601–602, Morgan Kaufmann.

Guarino, N., Masolo, C., Vetere, G., 1999. OntoSeek: content-based access to the web. IEEE Intelligent Systems 14 (3), 70–80.

Lepine, P., Aussenac-Gilles, N., 1994. Modélisation de la résolution de problèmes: comparaison expérimentale de KADS et MACAO. Actes des 5èmes Journées d'Acquisition des Connaissances, Strasbourg, mars 1994.

McDermott, J., 1988. Preliminary steps toward a taxonomy of problem-solving methods. in: Marcus, S. (Ed.), Automating knowledge Acquisition for Expert Systems. Kluwer Academic Publishers.

Musen M., The knowledge acquisition workshops: a remarkable convergence of ideas, International Journal of Human–Computer Studies, http://dx.doi.org/10.1016/j.ijhcs.2012.10.011, this issue.

Nelson, T.H., 1965. Complex information processing: a file structure for the complex, the changing and the indeterminate. In: *Proceedings of the 1965 20th National Conference of the Association for Computing Machinery,* pp. 84–100, New York, NY, USA. ACM Press.

Newell, A., 1980. The Knowledge Level. Presidential Address, American Association for Artificial Intelligence. AAAI', 80. Stanford University.

Norman, D., 1986. A. et. in: Draper, S.W. (Ed.), User Centered System Design. Lawrence Erlbaum Associates Publishers.

Reynaud, C., Tort, F., 1994. Connaissances du domaine d'un SBC et ontologies: discussion. Actes des 5èmes Journées d'Acquisition des Connaissances, Strasbourg, mars 1994.

Reynaud, C., Tort, F., 1996. Use of expertise ontologies in the knowledge engineering process. Journal: IEEE Transactions on Applications and Industry, 106–109.

Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W., Wielinga, B., 1999. Knowledge Engineering and Management, The CommonKADS Methodology. MIT PressMIT Press ISBN-10: 0-262-19300-0.

Tudorache T., N.F. Noy, S.W. Tu, M.A. Musen. 2008, Supporting collaborative ontology development in Protégé. In: Proceedings of

Seventh International Semantic Web Conference, Karlsruhe, Germany, Springer.

Vogel, C., 1988. Le génie Cognitif, 180. Masson, Paris pages.

Wielinga, B. Reflections on 25+ years of knowledge acquisition. International Journal of Human–Computer Studies, http://dx.doi.org/10.1016/j.ijhcs.2012.10.007, this issue.

Zacklad, M., Rousseaux, F., 1996. Modelling co-operation in the design of knowledge production systems: the MadeIn'Coop method. Computer Supported Cooperative Work: The Journal of Collaborative Computing 5, 133–154.

Zacklad, M., 1993. Les SBC vus comme des Systèmes Interactifs d'Aide à la Résolution de Problème. Actes des 4emes Journées d'Acquisition des Connaissances du PRC-IA, Saint-Raphaël]l, 31 Mars-2 Avril 1993.

Zacklad M., 1999. Introduction aux ontologies sémiotiques dans le Web Socio Sémantique, in actes de la conférence, In: Jaulent, M.-C. *16èmes journées francophones d'Ingénierie des Connaissances*, Grenoble: PUG.