

ERCIM “Alain Bensoussan” Fellowship Scientific Report

Fellow: Giorgos Flouris
Visited Location: CNR, Pisa, Italy
Duration of Visit: 9 months (04/09/06 – 03/06/07)

I - Scientific activity

The research activities performed during my ERCIM “Alain Bensoussan” Fellowship period can be roughly divided in two general research areas: the problem of change (belief revision and ontology evolution) and the problem of digital preservation.

The problem of change refers to the general problem of updating a corpus of knowledge in the light of new information. The work performed in this area was largely a continuation of the work performed during my PhD. A significant portion of this work is as yet unpublished, however its results are expected to be exploited and published in the upcoming months. In my dissertation, I showed that a popular belief change methodology (namely, the AGM theory) cannot be used in many logics useful for practical applications, including certain Description Logics; during my fellowship period, I worked on a weakening of the AGM theory, so as to be applicable to such logics. Moreover, I classified change operations in three different levels [5], a work which turned out to be related to the preservation problem as well. Other relevant activities include a literature survey regarding ontology change [6], [12], some results related to the updating of RDF ontologies [10] and a comprehensive summary of my PhD thesis [2].

The second general research area I worked on during my fellowship period is the problem of digital preservation, which refers to the problem of rendering a digital object readable and comprehensible in the future, despite the possible technology shifts and other changes that may have occurred in the interim, causing, e.g., software and hardware incompatibilities. This work was directly related to the CASPAR project and the OAIS model, but was, in large part, focused on the theoretical aspects of the problem. We identified a number of requirements for a theoretical model of digital preservation [3] and proposed some preliminary ideas towards a formalization of the problem [8], [11]; in these works it is argued, among other things, that the problem of digital preservation has strong connections with the problem of change. Another relevant research direction identified the notion of a “knowledge gap”, which refers to the information required in order to understand an otherwise non-intelligible digital object, and described how to model and manipulate this kind of information [4]. Finally, part of the work on digital preservation focused on practical interoperability aspects associated to the problem [9] and other technical issues related to the CASPAR project.

Apart from these main research activities, I also performed some work on e-learning [1], participated in certain events (see section III below for details) and was involved in a number of organizational activities (participation in Program Committees etc), the most important one being the organization, as a co-chair, of the International Workshop on Ontology Dynamics (IWOD-07) [7] to be held as part of the 4th European Semantic Web Conference (ESWC-07), in Innsbruck, Austria, on 7 June, 2007.

II- Publication(s) during your fellowship

During my fellowship the following papers were submitted and/or accepted; some of them involve work partly developed before my fellowship started.

Journal Publications

- [1] Yannis Tzitzikas, Vassilis Christophides, Giorgos Flouris, Dimitris Kotzinos, Hannu Markkanen, Dimitris Plexousakis, Nicolas Spyrtos. Emergent Knowledge Artifacts for Supporting Trialogical E-Learning. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 2(3), pages 16-38, 2007.

Abstract. This article elaborates on scenarios for collaborative knowledge creation in the spirit of the trialogical learning paradigm. According to these scenarios, the group knowledge base is formed by combining the knowledge bases of the participants, according to various methods. The provision of flexible methods for defining various aspects of the group knowledge is expected to enhance synergy in the knowledge creation process and could lead to the development of tools that overcome the inelasticities of the current knowledge creation practices. Subsequently, these scenarios are projected to various knowledge representation frameworks and for each one of them, we analyze and discuss related techniques and identify issues that are worth further research.

- [2] Giorgos Flouris. On Belief Change in Ontology Evolution. *AI Communications Journal (AI-Com)*, 19(4), pages 395-397, PhD Thesis Summary, 2006.

Abstract. We study the feasibility of applying the dominating paradigm of belief revision, the AGM theory, to logics outside its original scope, with special emphasis given on logics used for ontological representation in the Semantic Web. Such an application would allow determining whether a proposed change operator (e.g., an ontology evolution algorithm) behaves rationally or not. We give a number of interesting theoretical results related to the (generalized) AGM theory and show that our work can find important applications in dynamic environments employing non-classical logical formalisms. We focus on a particular such environment, namely ontology evolution in the Semantic Web, and show how our work can be used to provide ontology evolution researchers with powerful formal tools from the research area of belief revision.

Peer-Reviewed Conference and Workshop Publications

- [3] Giorgos Flouris, Carlo Meghini. Terminology and Wish List for a Formal Theory of Preservation. To appear in *Proceedings of PV-2007: Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data (PV-07)*, 2007.

Extended Abstract. One of the most difficult problems being faced by modern archivists is the rapid obsolescence of large volumes of digital (especially “born-digital”) information. This problem is being addressed in the research area of information preservation. Preserving digital information is a very hard problem, not fully understood to date; in particular, there is no commonly accepted formal model to describe it or a formal description of the required properties of a good solution. This paper presents some ideas towards filling this gap by focusing on the theoretical dimensions of the

problem and proposing some preliminary formal definitions and requirements for a theory of preservation.

We begin with a general analysis which identifies the need to define three different types of preservation. The first type, bit preservation, refers to the ability to read a particular sequence of bits; this can be achieved using error correction techniques, backups, RAID or mirrored disks, media refreshment and other technologies. The second type, data preservation, refers to the ability to render the digital object and produce a meaningful output from the data. This preservation type is the focus of most current approaches to the problem. The third type, information preservation, refers to the ability to understand the rendered file, i.e., to be able to understand its content by understanding the terms, concepts or other information that appears in it, by placing it in its correct context etc. This is the toughest type of preservation, and is often ignored by existing preservation approaches. We argue that a complete preservation model should handle all three preservation types.

We continue by presenting a number of examples which show that a central concept for preservation is that of the “meaning” of a digital object. Our analysis shows that any digital object is given meaning using certain assumptions related to its format, context, terminology and other commonsense and background knowledge, most of which is often implicit. We provide convincing argumentation that this background information can and should be captured using some kind of logical formalism (not necessarily the same for all digital objects) plus a logical theory, expressed in terms of this formalism, which captures the community’s background knowledge. This structure is called the underlying community knowledge. Thus, each digital object is associated with a certain underlying community knowledge which provides its meaning.

Regarding the digital object, we argue that it is not usually necessary (or possible) to preserve the entire information carried by it; instead, we could isolate and preserve the object’s most “useful” or “important” information. Determining the information worth preserving for the object at hand is not an easy task; it depends on the object type, its content, legal issues as well as on the needs of the creator and the reader of the information. A great aid in this task is provided by preservation models, such as the OAIS, which we embrace in this work. The role of such a model in this respect is to provide a methodological framework and a “best practices” approach towards the aim of determining the most important pieces of information contained in a digital object. Based on these fundamental notions, we formally define concepts like preservation policy, preservation system, successful preservation, emulation, migration etc, as well as OAIS-related notions, such as the consumer, the producer, the designated community etc. We define desired properties for a preservation system and provide some discussion on the issues related to the evolution of the underlying community knowledge and the effects of such an evolution on the associated objects. Finally, we show that, under certain conditions, this evolution can be modelled using techniques from the well-established research fields of ontology evolution and belief revision.

- [4] Yannis Tzitzikas, Giorgos Flouris. Mind the (Intelligibility) Gap. To appear in Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL-07), 2007.

Abstract. Intelligibility, evolution and emulation are some of the key notions for digital information preservation. In this paper we define formally these notions on the basis of modules and inter-module dependencies. Subsequently, we discuss how we can handle the evolution of modules and dependencies. This work can be exploited for building advanced preservation information systems and registries.

- [5] Giorgos Flouris. On the Evolution of Ontological Signatures. In Proceedings of the Workshop on Ontology Evolution (OnE-07), 2007.

Abstract. During ontology evolution, we are often faced with operations requiring the addition/removal of some ontological element (e.g., a concept) to/from the signature. Such operations deal with the ontological signature and are fundamentally different from operations that deal with the axiomatic part of the ontology, because they don't affect our knowledge on the domain but the non-logical symbols of the logic used to represent our knowledge on the domain. The consequences of this observation have been generally disregarded in the relevant literature. This paper attempts to fill this gap by introducing the concept of "change levels" and discussing the issues emerging from the different nature of the two types of operations. Furthermore, two alternative formalizations are described, which allow both types of operations to be represented at the same level, and, consequently, be considered of the same type.

- [6] Giorgos Flouris, Dimitris Plexousakis, Grigoris Antoniou. A Classification of Ontology Change. In the Poster Session of the 3rd Italian Semantic Web Workshop, Semantic Web Applications and Perspectives (SWAP-06), 2006.

Abstract. The problem of modifying an ontology in response to a certain need for change is a complex and multifaceted one, being addressed by several different, but closely related and often overlapping research disciplines. Unfortunately, the boundaries of each such discipline are not clear, as certain terms are often used with different meanings in the relevant literature. The purpose of this paper is to identify the exact relationships, connections and overlaps between these research areas and determine the boundaries of each field, by performing a broad review of the relevant literature.

Edited Books and Proceedings

- [7] Giorgos Flouris, Mathieu d'Aquin (editors). Proceedings of the International Workshop on Ontology Dynamics (IWOD-07). 2007.

Invited Talks and Publications

- [8] Giorgos Flouris, Carlo Meghini. Steps Towards a Theory of Information Preservation. In International Workshop on Database Preservation (PresDB-07), Invited Talk, 2007.

Abstract. The rapid obsolescence of large volumes of digital data is one of the most difficult problems faced by modern archivists. This problem (information preservation) is not fully understood to date; one of the major gaps in related research is that there is no general agreement on a formal model to describe the problem or on a formal description of the required properties of a good solution to the problem. This work's aim is to fill this gap by developing a formal, mathematical, logic-based description of preservation as a scientific discipline, to the end of deriving a methodology resting on solid theoretical ground. Thus, this work is not about how to perform preservation, but about how to describe formally what it means to perform preservation.

We will present a number of preliminary results towards this aim, including the identification of three types of preservation, some connections with the related fields of belief change (belief revision) and ontology evolution, and some desirable properties of a formal description of preservation. Our study will show that it is not usually necessary (or possible) to preserve the whole digital object (e.g., a database) but a set of properties

(questions and their answers) related to the object, and will verify the need to separate the data from its environment (i.e., the external information which gives meaning to the data). Both issues are considered in our proposal towards a formal description which handles both the static and the dynamic aspects of the problem.

Other Articles (to be submitted or under review)

- [9] Fabio Nanni, Giorgos Flouris. Issues Related to a Soft Transition of Existing Archives to a Preservation Environment. To be submitted.

Abstract. This document addresses the problem of ruling a soft transition of an existing archive to an externally managed preservation environment. When such a transition must take place, the archive management has to decide whether to hand over the complete custody of its data holdings and/or its procedures to the preservation system, or to retain control of any of these, and, if so, to what extent. We are motivated by the admittance that most archives will probably be quite reluctant in abandoning their data, as well as their well-established procedures, protocols and interfaces, in favor of something external to the archive. This admittance raises a number of issues related to how to make the transition of the archives into the preservation system as smooth as possible and how to facilitate the interaction between the two systems.

- [10] Georgios Konstantinidis, Giorgos Flouris, Grigoris Antoniou, Vassilis Christophides. On the Use of Ordering Relations for RDF Updating. Under Review for the 6th International Semantic Web Conference.

Abstract. We introduce a formal framework to handle the problem of updating an RDF Ontologies. Inspired by similar belief revision notions, we formalize two basic requirements for a correct update, namely success and validity; among the many different update results satisfying these requirements, we select the best one, defined as the one that is “closest” to our original knowledge, following the well-known Principle of Minimal Change. The above notion of closeness is determined by an operation-independent ordering relation, rather than a set of per-operation heuristics, which allows our approach to exhibit a more consistent behavior with respect to the consequences of a change operation; in addition, this behavior can be tuned by changing the ordering relation. Other merits of our approach include its ability to deal with implicit knowledge, its formal character and deterministic behavior as well as its ability to, transparently, deal with any imaginable change operation or set of operations.

- [11] Giorgos Flouris, Carlo Meghini. Some Preliminary Ideas Towards a Theory of Digital Preservation. Under Review for the 1st International Workshop on Digital Libraries Foundations.

Abstract. The problem of digital preservation is one of the most challenging research problems faced by the community of digital libraries today, receiving growing interest by researchers and practitioners alike. One of the major gaps in the related research is the lack of a general agreement on a formal model to describe the problem or on a formal description of the required properties of a good solution to the problem. This work aims to fill this gap by presenting a number of ideas towards a formal, mathematical, logic-based description of preservation as a scientific discipline, to the end of deriving a methodology resting on solid theoretical grounds. We will present and justify a number of desired properties of such a formalism and introduce a model that

handles the static aspects of the problem; some ideas related to the dynamics of preservation will be presented as well.

- [12] Giorgos Flouris, Dimitris Manakanatas, Haridimos Kondylakis, Dimitris Plexousakis, Grigoris Antoniou. *Ontology Change: Classification and Survey*. Under Review for the Knowledge Engineering Review journal (KER).

Abstract. Ontologies play a key role in the advent of the Semantic Web. An important problem when dealing with ontologies is the modification of an existing ontology in response to a certain need for change. This problem is a complex and multifaceted one, because it can take several different forms and includes several related subproblems, like heterogeneity resolution or keeping track of ontology versions. As a result, it is being addressed by several different, but closely related and often overlapping research disciplines. Unfortunately, the boundaries of each such discipline are not clear, as the same term is often used with different meanings in the relevant literature, causing a certain amount of confusion. The purpose of this paper is to identify the exact relationships between these research areas and to determine the boundaries of each field, by performing a broad review of the relevant literature.

III -Attended Seminars, Workshops, and Conferences

During my fellowship period, I attended the following events:

- The Workshop on Ontology Evolution (OnE-07), held as part of the 10th International Conference on Business Information Systems (BIS-07), in Poznan, Poland, on 25-27 April, 2007.
- The International Workshop on Database Preservation, held in Edinburgh, Scotland, on 23 March, 2007.
- The 3rd Italian Semantic Web Workshop, Semantic Web Applications and Perspectives (SWAP-06), held in Pisa, Italy, on 18-20 December, 2006.