Fourth International Conference on Semantics, Knowledge and Grid

A Web of Active Knowledge

Joe Geldart[#1], Stephen Cummins[#2], William Song[#3]

[#]Department of Computer Science, Durham University
University Science Laboratories, South Road, Durham, UK, DH1 3LE

[1]j.r.c.geldart@dur.ac.uk
[2]s.a.cummins@dur.ac.uk
[3]w.w.song@dur.ac.uk

Abstract— **This paper outlines a new model for the Web inspired by recent research into cognitive science. It is argued that the trend towards decentralisation requires a change from a typical client-server meta-model to one based on the idea of communicating information systems. It is further argued that to accomplish this, one must accept the idea of information as dynamic and embodied. The meta-model described is then applied to a typical existing information systems based on the Semantic Web 'wedding cake'.**

I. Introduction

The Web is what is termed a disruptive technology. That is, a technology which causes large changes to the way in which society operates. This disruption has taken the form of increased openness, and accessibility, of information to all its users. In today's world, anyone can be a publisher and anything can be shared. The Web enables its users to communicate with an ease which would have been difficult to imagine twenty years ago, decentralising and globalising knowledge dissemination.

This levelling effect of everyone being both consumer and publisher is now driving a new wave of user-generated content; everyone has something to say, and a way of saying it. However, little work has been done to question the underlying technologies of the Web. With a massive increase in traffic, and an accelerating trend towards user participation, we feel it is the right time to ask whether there can be a better model for the Web and what this may look like.

Simultaneous to the development of the Web, research into cognitive science and related disciplines have given us insights into the way in which humans conceptualise and communicate knowledge about the world. Important ideas such as embodiment, and forms of reasoning called blending, have provided insights into the nature of human semantics and communication. Embodiment in particular provides an interesting route to solving the problems that disagreement presents.

This paper outlines a new, cognitively inspired model for the Web based upon the notion of information processing agents communicating active information. The main contributions of the paper are:

- The identification of fundamental problem of disagreement and the active nature of information (section 2.)
- The description of a novel, two-level meta-model for the Web based on the idea of the communication of active information between information processing systems (section 3.)
- The application of this model to the description of typical, existing systems along with an explanation of the process of adapting these systems to our meta-model (section 4.)

We finish with a brief survey of the related work in this area, especially focusing on the inspirations for our model (section 5.)

II. A Web of Disagreement

No matter how hard one may endeavour to convince someone of something, there will always be those who disagree. It is a desideratum for open systems, such as the Semantic Web [1], that they be robust in the face of conflicting information.

Current trends in the Web appear to be towards an increase in the decentralisation and openness of information creation. Growing technologies such as cooperative tagging systems allow their users to annotate resources with short, pithy tags which are relevant to themselves without enforcing a universal vocabulary. This allows users to build up a personal system that facilitates information retrieval as well as allowing searches across all users' tags. However, the lack of central control, whilst useful in many cases, leads to a situation where disagreement becomes rife and there is a lack of quality control.

It is our contention that, in order to evolve the underlying model of the Web to cope with disagreement and achieve a useful and stable decentralised information system, we must explicitly model the way in which information itself evolves and flows between users. We propose the development of a new model of the Web that explicitly focuses on the way humans communicate and process knowledge, rather than on how machines do. We call this model the Active Web, as its key aspect is the treatment of information as an active entity, which may alter other information [2].

III. An Active Information System Meta-model

The current models of the Web are very passive and static things. By contrast, humans are active and dynamic. All the information that is on the Web is a product of human action in some form. Now, human conception is not a static thing; we are not born with everything we are ever going to know. We learn, we adapt, we make mistakes in our beliefs that we then

978-0-7695-3401-5/08 $25.00 © 2008 IEEE
DOI 10.1109/SKG.2008.103

correct. We perform this weighing of evidence and subsequent adaptation of belief without noticing.

It should be clear from this that the assimilation of information, its understanding and subsequent dissemination, can be seen as a form of process. We learn by acting. We communicate by acting. Our use of the Web is just a particular form of action, allowing us to find parcelled snippets of another's thoughts. There is no inherent semantics to the information on the Web, just the meanings we acquire through our readings. It is important then that the meanings information systems acquire are based on the evidence they experience [3][4].

This then is what we propose as the future of the Web; accepting the fundamental role that process, particularly communication, plays in defining meaning. The Web exists to enable communication. This entails a number of important changes in perspective. Rather than treating information on the Web as having meaning in and of itself, it only gains its meaning through the interactions of its users, be they machine or human, and its embodiment in the world that the users occupy. Equally, the desire for openness and decentralization entails that we accept disagreement and provide mechanisms to deal with it that do not require us to discard information unnecessarily. Information is active in that it affects the understanding of other information.

Rather than a fundamental distinction between producers (the servers) and consumers (the clients), we should treat production as merely as one outcome of the consumption process. We propose a Web of communicating agents, with no a priori distinctions between them, all (at least notionally) communicating on the same footing using symmetric protocols. This proposal, of a Web of active and communicated information produced and consumed by active agents acting within the dynamic real world, we term the Active Web [2].

The Active Web is conceived of as consisting of two separate, yet related levels. The first level is that of inter-agent communications. Imagine a typical university, such as Durham. A typical grouping is a department. Each department has associated with it a number of pieces of information that define its collective knowledge. These include things like research notes and course materials but also student information and administrative forms. This knowledge is the shared knowledge of its members, which forms the basis for the communication of information pertaining to the activities of the department and its links with other departments. We term such a collection of shared knowledge, a cloud. It has a gradual boundary, is dynamic and constantly shifting and may mingle with other clouds as they come near. A cloud is the fine glue that connects a collection of agents, the matters being communicated and the agents themselves (as each agent has knowledge regarding the other agents connected by the cloud.) We see the Active Web as being glued together by these clouds of knowledge. The exchange of knowledge between clouds causes them to grow. This increases the intersection that may be used as the cultural basis of future communication.

The communication model for the Active Web is based upon the notion of active agents transmitting self-contained bundles of information with a logic with an evidence-based semantics and instructions for combining such bundles.

We can consider a document as an example of a conceptual space — that is, an encapsulated bundle of inter-related data (information) together with the structure, models and inference rules of the language in which the information is expressed (the space's logic) — which is sufficient to communicate a suite of ideas to a community of agents. Such conceptual spaces we shall call *communicons*. A minimal communicon contains just enough information to convey the space once it is blended with the receiving agents pre-existing spaces. Additional information over this is called redundancy.

A collection of spaces able to be rendered as communicons for a given agent is what we call a knowledge cloud. The intuition is that knowledge, as justified belief, is information able to be connected to some agent's existing beliefs (the justification of a knowledge cloud, or the rheme of the cloud). The product of the spaces in a knowledge cloud is called the cloud's topic.

We propose to model these conceptual spaces, communicons, knowledge, and the connections between them as institutions [5] — a construct from category theory. The notion of institutions has already been applied to formal specification languages [6] as well as to cognitive science [7]. We propose to use computational models of these conceptual spaces as institutions to provide a formal, yet practical, model of the spread and growth of knowledge in the Active Web.

As a concrete example, consider the role of a departmental secretary. One may imagine the secretary as a facilitator agent; distributing information and helping other agents forge new connections. As the secretary answers questions, their own knowledge grows as they integrate and pass on information. By trying to form new blends by integrating existing conceptual spaces, novel knowledge may be formed, causing growth in the secretary's cloud and thence those of the other agents.

The second level is that of the structure of the agents themselves. Each agent is a self-contained, active information system, taking incoming information, processing it and producing an effect on the world. Rather than assuming that all information systems are the same, instead we propose a meta-model which is inspired by research into neurology and cognitive science. We call systems that use this meta-model, *cogs*. Each cog consists of a stack of four tiers: an embodiment tier, a primary sensory/motor tier, a secondary sensory/motor tier and an associative tier. All the tiers, apart from the associative tier, are divided into two modalities; sense and action (the 'motor' modality). Information enters the system through the sensory modality and the system acts upon this information through the motor modality.

The embodiment tier connects the cog to other cogs and to input and output systems in the real world. The sensory modality of the primary tier takes raw information from the embodiment tier and extracts relevant data features from it. This is then passed to the secondary tier, which is responsible for organising these features into concepts; in other words, turning data into information. The associative tier processes these concepts, in light of past experience, and chooses abstract action plans (operators) to invoke. These plans are decomposed into component action features by the motor modality of the secondary tier, and these are thence transformed into execution requests by the primary tier. The motor modality of the embodiment tier then manages the concrete execution of these requests by communicating with other cogs, or by carrying out actions upon the world. A pictorial summary of this architecture is given in Fig. 3, also showing some possible connections between cogs and objects in the external environment.

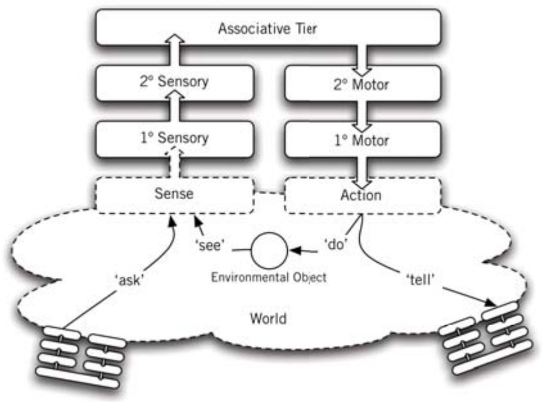

Fig. 2. The Cog Meta-model.

This architecture bears some similarity to the subsumption architecture [8]. Each tier can embellish the information it receives in order to account for its own concerns, such as pragmatic impediments to execution. By providing a common meta-model for active information systems, we intend that common protocols may be devised for connecting the tiers. This will allow complex, novel information systems to be built out of pre-existing parts.

In this model the processing and transformational role of the component systems is made explicit. Cogs do not merely passively consume information held in relatively static documents, but rather play an active and essential role in the creation, modification and dissemination of the Web's information.

IV. INTEGRATING EXISTING SYSTEMS

An advantage of the Cog meta-model is that it not only guides the development of new, active information systems but that it can also incorporate existing information systems; in other words, the notion of cogs is an evolution, rather than a revolution in the design of information systems. To demonstrate this, we now describe how systems using existing models may be transformed into the common meta-model of cogs. We focus on one existing sort of system: a Semantic Web application based on the 'wedding cake'. This reformulation makes explicit the whole processing loop, from information acquisition through processing to output. Such an explicit treatment of the activeness of the information system's embedding in the world is often missing from existing architectures for these systems. It is hoped that by having a single meta-model capable of describing different, existing styles of system (as well as many systems as yet not thought of), the notion of a cog will help integrate the shallow and the deep Webs.

Recently, data browsers such as Tabulator [9] have come to prominence as a means of exploring the links between data in the Giant Global Graph. These systems are based on the 'wedding cake' stack of technologies. Whilst the use of this stack is fairly well accepted, the architecture does not include much detail on how its layers are connected. We will see how the Cog meta-model makes such connections more explicit.

A typical data browser gathers information as the user browses the Web. The user can then switch to a data view which allows them to pick nodes and facets to apply in order to produce summaries which can be rendered (often using XSLT.) The system may apply rules based on the ontological classes in the data graph in order to derive new relationships from old.

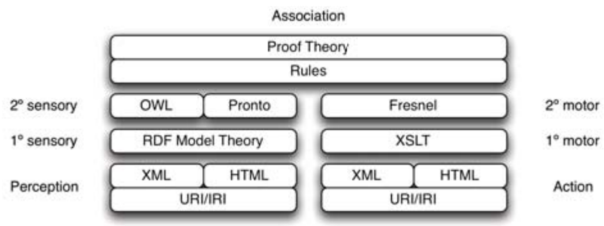

Fig. 5. A wedding cake system as a cog.

This data browser architecture is a straightforward example of the Cog meta-model (Fig. 5). The data browser is embodied via its connection to a web-browser through the standard notion of a URI. The input to the sensory modality consists of browsed pages and requests to view the metadata of a particular object. RDF information is extracted from the pages, and the scraped RDF triples are the relevant features of the input. An ontology language, such as OWL, may then be applied to classify the nodes and rules can be used to derive new information. A facet description language like Fresnel [10] forms abstract output descriptions based on the classes of the nodes which may be rendered to a more concrete form using XSLT which can then be delivered to the user.

In contrast to the existing stack approach, the use of the Cog meta-model makes explicit the flow of information from input to output, and the way that the system as a whole is embodied in the world. By making these significant details explicit, it is easier to see how such systems may inter-relate with other systems, which may not use the same technologies or be considered to be a Semantic Web application.

V. RELATED WORK

In the information systems community, the most similar meta-model to our one is that developed in [11]. They propose a model, called the Global System Model, which treats an information system as consisting of information sources linked to an information processor. There can be many sources of input to the system. The processor then consists of three phases: knowledge importation and extraction, knowledge abstraction and knowledge integration. Conflicts are resolved in the integration phase (much as they must be in the associative tier in the Cog meta-model) and users may browse the concepts using a hypertext browser.

The main differences between our meta-model and the Global System Model is the explicit separation of the sensory and motor modalities and the changes caused by the different underlying philosophy of embodiment. Unlike the Global System Model, users do not peer in at the top after integration is done but rather ask questions which go through the same phases of processing as beliefs do. We explicitly discourage the stepping out of the world; the embodiment tier should mediate all the system's interactions with other systems. The use of two modalities will allow us to model the way in which information reacts with other information, is changed and then spread to other systems.

The inspiration for the tiered meta-model comes from research into the modular processing structure of the brain. We believe that machines should adapt to people, rather than people to machines and so feel that generic information processing should, as far as possible, follow a cognitively inspired architecture. [12] described a (symbolic) architecture for mental processing based upon chains of processors which successively transform information using innate and learnt associative rules. Pairs of processors communicate over well-defined interfaces which restrict what information is accessible to each other. This architecture is in contrast to the modular architecture of [13] where the modules are entirely black boxes. The notion of a limited interface is intended to provide a balance between non-modularity and Fodorian black box modularity. It is at this balance point that we intend our own work to sit.

The Cog meta-model makes no commitment to the use of a single processing or representation style across all systems, or even throughout a single information system; each tier may be composed of multiple modules that may be implemented in many ways. It is expected, indeed hoped, that technologies such as symbolic, soft and sub-symbolic computing may be combined where appropriate to form powerful hybrid information systems. We believe our work is the first to attempt to tackle the ideas of embodiment, information flow and the epistemology of communicating systems in a single, technically feasible model.

VI. Conclusions

In this paper, we have outlined a new model for Web systems based on communications between active information systems conforming to a particular meta-model. The fundamental aspects of our model are based on the philosophical position that the Web is based on the human need to communicate. As such, we've taken research into cognitive science, both at the neurological and functional levels, as our starting point.

Our model is divided into two levels, communication and processing, as a simplifying assumption. The two levels are linked by the notion of embodiment. The philosophical position of embodiment contends that the meaning of information in an information system comes from how that system is connected to the world. We believe that by modelling both the processing and communication of knowledge, we may build smarter systems. This is not just in the AI sense, but also in the technological. Possible applications of this include adaptive telecommunications architectures based on automated analysis of community structures and dynamic caching of data to minimise bandwidth consumption.

We have chosen to focus, for the moment, on the development of the Cog meta-model. This is primarily because of the necessary dependency relationship between the two levels of this programme; without the Cogs, we cannot build the communications layer. However, conversely, one needs to understand the nature of the communications between Cogs in order to build the Cogs themselves. We have chosen the ideas of information flow, and the formal realisation of conceptual blend theory, as the starting point for our research into the communications model of the Active Web. We intend to develop the two levels of our model in parallel, allowing discoveries in each to inform the other.

References

[1] Berners-Lee, T., Lassila, O. and Hendler, J. (2001) 'The Semantic Web.' *Scientific American*, May 2001.

[2] Geldart, J. and Song, W. (2008) 'Active: An Essential Nature for the (Future) Web?' In De Roure, D. and Hall, W. (eds.), *Proceedings of the First International Workshop on Understanding Web Evolution (WebEvolve2008): A prerequisite for Web Science*. Web Science Research Initiative: Southampton, UK. ISBN 978 085432885 7.

[3] Bussler, C. (2008) 'Is Semantic Web Technology Taking the Wrong Turn?' *IEEE Internet Computing*, 12, 75–79.

[4] Schoop, M., de Moor, A. and Dietz, J. (2006) 'The Pragmatic Web: A Manifesto.' *Communications of the ACM*, 49, 75–76.

[5] Goguen, J. and Burstall, R. (1992) 'Institutions: Abstract Model Theory for Specification and Programming.' *Journal of the ACM*, 39, 95–146.

[6] Goguen, J. and Meseguer, J. (1986) 'Extensions and Foundations of Object-Oriented Programming.' *ACM SIGPLAN Notices*, 21, 153–162.

[7] Goguen, J. (1999) 'An Introduction to Algebraic Semiotics, with Application to User Interface Design.' In Nehaniv, C. (ed) *Computation for Metaphor, Analogy and Agents*, Springer-Verlag.

[8] Brooks, R. (1990) 'Elephants Don't Play Chess.' *Robotics and Autonomous Systems,* 6, 3–15.

[9] Hendler, J. (2001) 'Agents and the Semantic Web.' *IEEE Intelligent Systems*, 16, 30–37.

[10] Pietriga, E., Karger, D., Bizer, C. and Lee, R. (2006) 'Fresnel: A Browser-Independent Presentation Vocabulary for RDF.' In Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M. and Aroyo, L. (eds.) *The Semantic Web: ISCW 2006*, Lecture Notes in Computer Science 4273, Springer-Verlag.

[11] Freeman, M. and Layzell, P. (1994) 'A Meta-Model of Information Systems to Support Reverse Engineering.' *Information and Software Technology*, 36, 283–294.

[12] Jackendoff, R. (1997) *The Architecture of the Language Faculty*. MIT Press.

[13] Fodor, J. (1983) *The Modularity of Mind: An Essay on Faculty Psychology.* MIT Press: Cambridge, MA.