INFORMATION SPACES AS A BASIS FOR PERSONALISING THE SEMANTIC WEB

Ian Oliver
Nokia Research Center, Itämerenkatu 11-13, Helsinki, Finland

Keywords: Space based computing, Semantic Web, Mobile devices.

Abstract: The future of the Semantic Web lies not in the ubiquity, addressability and global sharing of information but rather in localised, information spaces and their interactions. These information spaces will be made at a much more personal level and not necessarily adhere to globally agreed semantics and structures but rely more upon ad hoc and evolving semantic structures.

1 INTRODUCTION

In this paper we position our vision of the continuation of the development or evolution of the Semantic Web (Berners-Lee et al., 2001). This is best visualised as the Giant Global Graph concept popularised by Tim Berners-Lee[1].

Most information however is not ubiquitous but personalised, hidden, private and interpreted locally - this information tends to be the personal, highly dynamic information that one stores about oneself: contact lists, friends, media files, 'my' current context, 'my' family, 'my' home etc and the interweaving and linking between these entities through ad hoc personal structures.

We elaborate on the ideas of ubiquitous information, the role of reasoning and knowledge, the location of the information with relation to its ubiquity through the concept of projections from the Giant Global Graph called spaces. We then describe an implementation of an environment supporting these ideas in a mobile and personal context as well as many of the issues that this directly brings up with regards to what are semantics and how information is going to be dealt with in this context.

In the following sections we outline our position and areas of research relating to notions of personalisation, Semantic Web, information and its meaing and semantics as well as our implementation.

[1] *http://en.wikipedia.org/wiki/Giant_Global_Graph*

2 PERSONALISATION AND SPACES

The Semantic Web is succeeding in relatively small-scale, specific situations which are restricted to a given domain. If we expand the notion of a domain in a more orthogonal sense to encompass personal level then this suggests that we have a notion of a 'Personal Semantic Web' in which one can organise their own information according to these principles. The advantages of a Semantic Web based approach is that certain structures, schemata and semantics can be fixed enabling some - and this is an important point, we should not (and can not?) try to attempt everything - meaningful communication, reasoning and interoperability to take place.

Mobile devices with various methods of connectivity which now constitute for many as being the primary gateway to the internet and also being a major storage point for much personal information (Idehen and Erling, 2008; Lassila and Adler, 2003). This is in addition to the normal range of personal computers and furthermore sensor devices plus 'internet' based providers. Combining these devices together and lately the applications and the information stored by those applications is a major challenge of interoperability (Tolk and Muguira, 2003; Turnitsa, 2005).

This is achieved through numerous, individual and personal *spaces* in which persons, groups of persons etc can place, share, interact and manipulate webs of information (Krummenacher, 2008; Khushraj et al., 2004) with their own locally agreed semantics without necessarily conforming to an unobtainable, global

Oliver I. (2009).
INFORMATION SPACES AS A BASIS FOR PERSONALISING THE SEMANTIC WEB.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Software Agents and Internet Computing*, pages 179-184
DOI: 10.5220/0002155901790184

whole. These spaces are *projections* of the 'Giant Global Graph' in which one can apply semantics and reasoning at a local level. A detailed survey of such space-based systems is given in (Nixon et al., 2008).

This approach we feel addresses at least two of the counter-arguments against the Semantic Web vision: feasibility and privacy by directly addressing notions of locality or ubiquity and ownership. Feasibility because we are changing the problem to address much smaller-scale structures through setting clear boundaries in terms of computation.

In order to apply reasoning and other manipulations (such as sharing) of that information we are required to construct processes which have access to that information - typically these are known as agents in the 'traditional' sense of the word although we tend towards the classification given in (Haag et al., 2003). Agents are either personal in that they perform tasks either directly decided upon by the user or autonomously for or on behalf of the user, monitor particular situations or reason/data-mine the existing information.

Personalisation is achieved through explicitly demarking a space in which information is stored and agents have access. Within each space information is organised according to the owner (or owners) of that space. For an agent to obtain entry to that space then it is made on the terms of that space. Similarly for two spaces to interact directly similar contracts must also hold. Interactions with spaces is described in section 3.

The kinds of information that are stored in a personal space vary but initially contacts lists, media files (links to media), personal information management data (calendars etc), email and other personal communication etc. This easily expands to information feeds such as those provided through RSS and even WWW interfaces, family or community information, social networking and so on. These kinds of information can be then further augumented by tagging, internal links and more sophisticated equivalence relationships such as might be seen between social networking contacts, contacts lists, calendars etc. In addition more static and thus more externalisable information can be stored or referenced in the same manner - such information might be census records, telephone directories or even cultural information (Hyvönen et al., 2008).

Of course there are issues regarding the interpretation of information and how the meaning or semantics is preserved across spaces and agents; this also includes deciding whether two independent structures actually represent the same piece of information and can be merged or coalesced. Furthermore issues regarding trust and security need to be addressed - we do not specifically discuss this problem in this paper.

3 INTERACTION AND SHARING WITH SPACES

Consider the follow scenarios. In figure 1 Alice interacts with her personal space - through 'agents'[2] running on a multitude of devices. This space contains a corpus of information A which through local reasoning and deductive closure algorithms - a feature of our spaces - provides her with the corpus R(A). Information is represented using Semantic Web standards, ie: RDF, RDFS, OWL, FOAF etc, rule sets in RuleML etc.

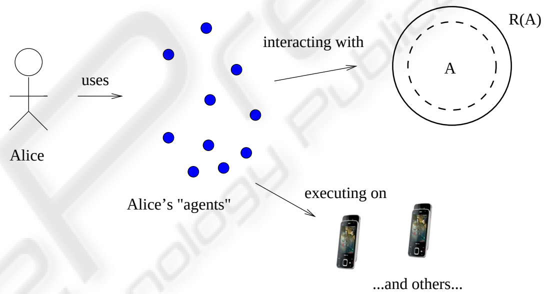

Figure 1: Alice's Agents, Devices, Spaces and Information.

In this case the boundary of R(A) is the limit of Alice's personal space. If Alice has two spaces or corpii of information she might bind these together to produce a much larger space. These individual corpii of information may be overlapping in terms of their content. Alice can interact simultaneously with many discrete spaces. These situations are visualised in figure 2; for simplicity we only show the spaces and closures. Here the total information available for a given space is the union of the deductive closure over all the individual corpii.

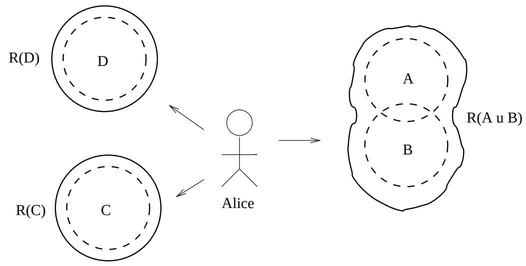

Figure 2: Alice and Alice's Spaces.

[2]Agent is rather a loaded word, but alternatives aren't numerous: executives, nodes, UIs, programs, etc

Alice can decide to break and reconfigure her current spaces into many smaller spaces. This may be made in any manner including removing all information and creating multiple individual smaller or even empty spaces to making a complete copies of the current space.

We now introduce Bob, who has a space of his own that is constructed from a single corpus of information. At this point we can say nothing about the relationship between the contents of the corpii A and B with C; they may potentially all contain the same information.

Interaction between Alice and Bob can be made in three different forms. The first is simple in that Bob only needs to give Alice's 'agents' access to his space as visualised in figure 3. Bob of course has his own ideas about privacy and grants Alice access to only a portion of his space. Alice has direct access to a subset of Bob's space - if she has write access then potentially this could have an effect on the space as a whole.

Privacy is asymmetric - it is on the sharer's terms only thus precluding the need for a globally agreed privacy mechanism. If Alice just so happens to be able to satisfy the criteria for accessing Bob's space then Alice is granted access at whatever level Bob's privacy mechanisms allow. Alice's own privacy mechanisms do not affect Bob's mechanisms and vice versa.

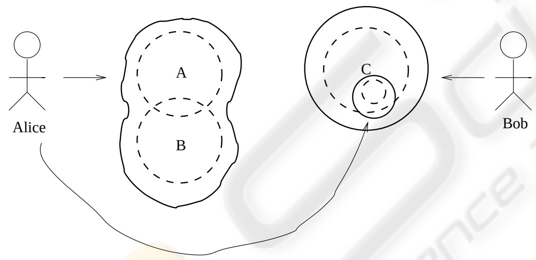

Figure 3: Alice Interaction with Bob's Space.

The second form of interaction is a variation of this: Bob to partitions off a given subspace to which Alice has access, then any changes can be kept local and the merge back controlled by Bob explicitly. In both of these cases access policy including trust mechanisms are local to Bob - there is no need for Alice to know about what mechanisms are in place.

The fist two forms we expect to be the most common methods of interaction, the third offers a set of different possibilities based around the merging of the spaces. This is more complicated as Alice and Bob must both agree to the merge (fig. 4) both in terms of personally agreeing through their trust mechanisms (might just be personal trust) but also through the shared semantics of the information.

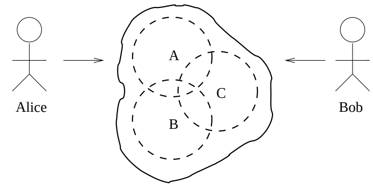

Figure 4: Alice's and Bob's (Merged) Space.

This then constitutes how spaces are related but has not yet addressed certain specific ideas about information and the semantics of that information and these are discussed in section 4.

4 INFORMATION AND SEMANTIC ISSUES

We can classify the issues with this approach as:

- Non-monotonicity of Deductive Closure and Rules
- Graph Provenance
- Semantics of the Information
- Uncertainty, Incompleteness, Inconsistency and Undefinedness of Information

Given a single space s, the information contained in that space is $i(s)$, the rules $\rho(s)$ and the deductive closure calculated as $R(i(s), \rho(s))$. A merge of two spaces s_1 and s_2 results in a single space s_m where s_m is calculated as $R(i(s_1) \cup i(s_2), \delta(\rho(s_1), \rho(s_2))$. The function δ determines the set of rules to apply and is constructed from a mechanism which prioritises the rules somehow - the exact mechanism would of course vary but we envisage would be similar to those found in certain kinds of non-monotonic calculi, eg. (Mueller, 2006).

Graph provenance and the related semantics problem are the major problems in any Semantic Web/interoperability system as there, beyond pre-agreed addressing and strict, standardised ontologies, no definitive method for relating structures representing the same information together exists.

While information about typing etc is carried within the space, deeper *semantics* is not. Identification of larger meaningful structures such as RDF molecules (Ding et al., 2005) provide at this time the strongest basis for provenance analysis and addressing deeper semantically meaningful structures. In the most part we must rely upon local convention between any two merging or interacting spaces with the hope that this leads to a more global convention

(Afraz Jaffri, 2008; Shafiq et al., 2008).

The meaning of the information to Alice might be very different to the meaning ascribed by Bob (see also provenance above). The meaning or semantics of the information can only be made by the reader (Burcea et al., 2003); the writer of the information only gives hints through typing and tagging and other relationships to what the intended meaning might be.

This hints to the question how do we guarantee that two agents understand the information in the same way, our response to this is that we don't care; at least to the point where there is no internal mechanism for this (Kim and Anseo-Dong, 2002). As it is, the meaning of some ontological structure evolves over time and there appears to be no reasonable mechanism for communicating deeper level semantic structures. For true, reliable communication how many semantic levels are required is unknown. However, since that if access to a space is granted or a merge made then this is implicit agreement between the parties and particularly the readers over the intended semantics of the information.

In (Gärdenfors, 2000) is provided a detailed discussion of the style of *intentional semantics* (Kim and Anseo-Dong, 2002) we propose here that it is not justing typing but the whole construct of properties of an object - and then the scope over which we define an object - that must be taken into consideration when deciding how an agent interprets a given structure. This also applies when dealing with the aforementioned graph provenance issue.

To complicate matters further, the notion of semantics embodied within these kinds of information structures is little more than meta-data whereas one really needs to describe a further relationship into the 'real-world' (Smith, 1996). Solutions based upon lexical and semantic analysis through resources such as Wordnet(Fellbaum, 1998) appear to be the most promising here with regard to issues surrounding semantics similarity (Jan Wielemaker and Wielinga, 2003; Ruotsalo and Hyvönen, 2007). However, despite this we can (an no current computer system) **never** be sure that any two agents actually act upon their respective interpretations information in the same way.

Currently we are seeing two mechanisms for semantic agreement: the first is through standardisation of ontologies (W3C) and the second is through folksonomic evolution of initially personal and informal structures into ad hoc ontologies which become more concretised as social agreements form. Even in the strict, formal ontology development scenario evolution of the ontology takes place as usage changes and develops (Ruotsalo et al., 2008), however this is much slower than the folksonomic cases. A fairly common mechanism for agreeing on semantics is the upper ontology approach, but again this suffers from the problem of *many* upper ontologies (for example: (Hyvönen, 2008))- an interesting discussion of this is made in (Wheeler, 2004)

Finally the problem of uncertain, incomplete, inconsistent and undefined information requires a much more formal approach within the various agent and reasoning structures (Lassila, 2008). Within the Semantic Web we must further explore notions of undefinedness, modality, probability and non-monotonicity. This is left for future work though mechanisms are already present within our implementation for attaching and modality properties.

5 IMPLEMENTATION

Supporting these ideas we already have the computing and networking infrastructure. Our particular solution builds a space-based computing framework (Oliver and Honkola, 2008; Oliver et al., 2008) based upon the Piglet/Wilbur (Lassila, 2007) RDF++ engine (Lassila, 2002). The notion of space being constructed out of a number of individual, linked (totally routable) brokers. Interaction with the space is via a agents which may reside on any suitable device with suitable connectivity and computing capabilities; similarly the brokers. Figure 5 shows one possible configuration.

Agents may connect to one or more spaces at a time and to which spaces may vary over the lifetime of an agent. Mobility in this sense is provided in a 'pi-calculus' manner (Milner, 1999) in that links may move rather than the physical running process. Agents can save state and become 'mobile' when another agent restores that state. We have not addressed code mobility in the current implementation and this remains low priority at this moment.

Agents themselves are anonymous and independent of each other - there is no explicit control flow other than that provided through preconditions to agent actions. A coordination model based around expressing coordination structures as first-order entities is being investigated, however we are more focussed on collecting and reasoning over context. Control flow can be made outside of the space through agents explicitly sharing details of their external interfaces through the space - this has been successfully used in coordinating media streaming and storage devices via UPnP[3] and NoTA[4] for example.

[3]Universal Plug and Play http://upnp.org

[4]Network on Terminal Architecture http:// www.notaworld.org/

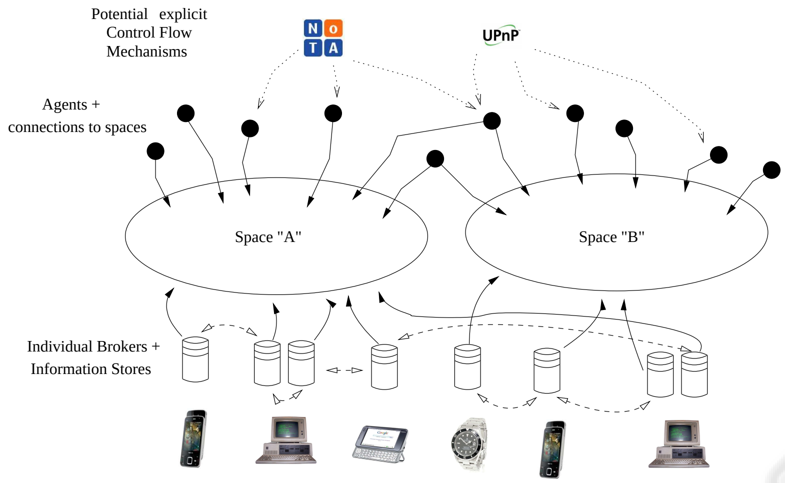

Figure 5: An Example Implementation Configuration.

The brokers each contain a corpus of information and when linked together to form a space distribute the information in an asymmetric manner - some information is not replicated because of computational resources, connectivity, storage reasons and even legal issues such as copyright. This distribution layer is also responsible for the efficient distributed computation of queries and marshalling the resultant responses such that calculating the deductive closure can be made. In this manner we can distribute complex computation away from the original agent over the whole space.

This distribution can be used in a number of ways and one of the most interesting for us is giving the user temporary access to a larger body of information which can be used to selectively enhance their current corpus of information, for example, a body of *linking information* based around the `owl:sameAs` construct (Passant, 2008). Other examples include personal access to more dynamic corpii such as news, weather or similar services.

We make no attempt to enforce consistency of the information rather letting the writer of the information have freedom to express what they want and leave the interpretation to the reader. The semantics of the information is merely intensional in that the writer provides 'clues' through typing, tagging and other means. Repair of information according to schemata or other criterion can be enforced within the space and for some kinds of information would even be desirable. If consistency of particular structures is required then this can be achieved through agent implementation and specific belief revision models.

6 CONCLUSIONS

In summary we believe that the Semantic Web will move from being a global information corpus to multiple, individual, linked and personal corpii. Semantics, reasoning and processing about the information will be localised and personalised within these corpii. As corpii are shared, linked, merged and split, certain schemata will coalesce and evolute into fixed or standard structures in an evolutionary form.

There are still questions regarding what precisely is semantics and how this is preserved across information structures but techniques do exist for reasoning about this and the related graph/information provenance problems and these are currently being implemented and trialled within the framework we have described.

REFERENCES

Afraz Jaffri, Hugh Glaser, I. M. (2008). Uri disambiguation in the context of linked data. In *Linked Data on the Web (LDOW 2008), Beijing, China*.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American*, 284(5):34–43.

Burcea, I., Petrovic, M., and Jacobsen, H.-A. (2003). I know what you mean: semantic issues in internet-scale publish/subscribe systems. In Cruz, I. F., Kashyap, V., Decker, S., and Eckstein, R., editors, *SWDB*, pages 51–62.

Ding, L., Finin, T., Peng, Y., Silva, P. P. D., and L, D. (2005). Tracking rdf graph provenance using rdf molecules. Technical report, 2005, Proceedings of the Fourth International Semantic Web Conference.

Fellbaum, C. (1998). Wordnet: An electronic lexical database.

Gärdenfors, P. (2000). *Conceptual Spaces: The geometry of thought*. The MIT Press. 0-262-57219-2.

Haag, S., Cummings, M., and McCubbrey, D. J. (2003). *Management Information Systems for the Information Age*. McGraw-Hill Companies. 4th edition. 978-0072819472.

Hyvönen, E. (2008). Finnonto-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi - visio ja sen toteus (finnonto model as the basis for a national semantic web infrastructure - vision and its implementation). In *Paper presented at the publication event of the Finnish Ontology Library Service ONKI*.

Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkarinen, P., Laitio, J., and Nyberg, K. (2008). Culturesampo – a collective memory of finnish cultural heritage on the semantic web 2.0.

Idehen, K. and Erling, O. (2008). Linked data spaces and data portability. In *Linked Data on the Web (LDOW 2008), Beijing, China*.

Jan Wielemaker, G. S. and Wielinga, B. (2003). Prolog-based infrastructure for rdf: performance and scalability. In *In: D. Fensel, K. Sycara and J. Mylopoulos (eds.) The Semantic Web - Proceedings ISWC'03, Sanibel Island, Florida. Lecture Notes in Computer Science, volume 2870, Springer-Verlag*, pages 644–658.

Khushraj, D., Lassila, O., and Finin, T. (22-26 Aug. 2004). stuples: semantic tuple spaces. *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, pages 268–277.

Kim, H.-G. and Anseo-Dong (2002). Pragmatics of the semantic web. In *Semantic Web Workshop. Hawaii, USA*.

Krummenacher, R. (2008). The smartest space of all: A global space of (machine-understandable) knowledge. In *1st Russian Conference on SmartSpaces, ruSmart 2008, St.Petersburg, Russia*.

Lassila, O. (2002). Taking the rdf model theory out for a spin. In *ISWC '02: Proceedings of the First International Semantic Web Conference on The Semantic Web*, pages 307–317, London, UK. Springer-Verlag.

Lassila, O. (2007). *Programming Semantic Web Applications: A Synthesis of Knowledge Representation and Semi-Structured Data*. PhD thesis, Helsinki University of Technology.

Lassila, O. (2008). Some personal thoughts on semantic web and non-symbolic ai. In *4th International Workshop on Uncertainty Reasoning for the Semantic Web*.

Lassila, O. and Adler, M. (2003). Semantic gadgets: Ubiquitous computing meets the semantic web. In Fensel, D., Hendler, J. A., Lieberman, H., and Wahlster, W., editors, *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, pages 363–376. MIT Press. 0-26206-232-1.

Milner, R. (1999). *Communicating and Mobile Systems: the Pi-Calculus*. Cambridge University Press. 0-521-65869-1.

Mueller, E. T. (2006). *Commonsense Reasoning*. Morgan Kaufmann. 0-12-369388-8.

Nixon, L. J. B., Simperl, E., Krummenacher, R., and Martin-Recuerda, F. (2008). Tuplespace-based computing for the semantic web: a survey of the state-of-the-art. *The Knowledge Engineering Review*, 23(2):181–212.

Oliver, I. and Honkola, J. (2008). Personal semantic web through a space based computing environment. In *Middleware for Semantic Web 08 at ICSC'08, Santa Clara, CA, USA*.

Oliver, I., Honkola, J., and Ziegler, J. (2008). Dynamic, localised space based semantic webs. In *WWW/Internet Conference, Freiburg, Germany*.

Passant, A. (2008). :me owl:sameas flickr:33669349@n00. In *Linked Data on the Web (LDOW 2008), Beijing, China*.

Ruotsalo, T. and Hyvönen, E. (2007). A method for determining ontology-based semantic relevance. In *Proceedings of the International Conference on Database and Expert Systems Applications DEXA 2007, Regensburg, Germany*. Springer.

Ruotsalo, T., Seppälä, K., Viljanen, K., Mäkelä, E., Kurki, J., Alm, O., Kauppinen, T., Tuominen, J., Frosterus, M., Sinkkilä, R., and Hyvönen, E. (2008). Ontology-based approach for interoperability of digital collections. *Signum*, (5).

Shafiq, O., Scharffe, F., Wutke, D., and del Valle, G. T. (2008). Resolving data heterogeneity issues in open distributed communication middleware. In *Proceedings of 3rd International Conference on Internet and Web Applications and Services (ICIW 2008), Athens, Greece*.

Smith, B. C. (1996). *On the Origin of Objects*. The MIT Press. 0-262-69209-0.

Tolk, A. and Muguira, J. A. (2003). The levels of conceptual interoperability model (lcim). In *Proceedings IEEE Fall Simulation Interoperability Workshop, IEEE CS Press*.

Turnitsa, C. D. (2005). Extending the levels of conceptual interoperability model. In *Proceedings IEEE Summer Computer Simulation Conference, IEEE CS Press*.

Wheeler, Q. D. (2004). Taxonomic triage and the poverty of phylogeny.