

Lessons for the future of Semantic Desktops learnt from 10 years of experience with the IDELIANCE Semantic Networks Manager

Jean Rohmer

CENTAI (Centre of Excellence For Information Analysis)
Thales Communications
146 Avenue de Valmy BP 82 92704 Colombes Cedex France
jean.rohmer@fr.thalesgroup.com

ABSTRACT

In 1993, we started to develop a tool to help people manage structured personal and business information. Initially named ORCCA (Online Resources for Corporate Citizens Action), it was fully implemented as a Semantic Networks Editor and Manager. Initially designed as a personal desktop tool, it evolved towards a collaborative server allowing small groups to share Semantic Networks. It has been marketed under the name of IDELIANCE, and has been used in significant business applications by a variety of users. During all these pioneer phases of design, implementation and usage of Ideliance, we learnt a lot of lessons which could be of interest for the success of the emerging Semantic Desktop domain. Not only did we implement practical tools for semantic networks management, but we also closely observed hundreds of people using Semantic Networks as a new way of writing and reading information. In this position paper, we propose a survey of the key points we encountered. For each of these points, we describe our experience, and we propose some guidelines for future Semantic Desktops design and usage.

Note : this paper is not a technical paper about Ideliance, or about its evolution towards the Semantic Web. It is a position paper which aims at expressing the general lessons we learned for the benefit of future designers and users.

HISTORY

History is important to position our work in its context so as to better reuse it in the current Semantic Web landscape.

In 1993, the so-called AI (Artificial Intelligence) *winter* was starting. In the 80's we had been deeply involved in the AI experience, as manager of a large products and services team (Groupe Bull Cediag), which achieved significant expert systems (like the SACHEM blast furnaces monitoring for ARCELOR) and innovative tools, like the first commercial implementation of constraint logical programming (CLP), CHARME. But a serious economical crisis made clear that these times were gone, and teams and projects disbanded. In the meantime, a new concept was emerging : « Knowledge Management » (KM), mainly boosted by new collaborative tools like Lotus Notes. Remember that, at this time the WEB simply did not exist in the Business and Industry world, even if the essential work had already been done by Tim Berners-Lee, and if Gopher and Mosaic were close to a reality.

Thus we decided to design a KM tool by recycling the simplest and most universal part of AI –its *representation* features- as opposed to its *automatic reasoning* features. And, for the sake of generality and simplicity, the Semantic Networks paradigm was chosen. This led us to develop a first version of a *personal semantic network editor* during the extended week-end of november 11, 1993. We then created a small company to develop this idea. We were involved in two important projects which influenced strongly the evolution of IDELIANCE : (1) The Mnemos project – an european EUREKA project with, among others, Aerospatiale (now EADS), Matra, French Atomic Energy Commission and NFT-Raufoss from Norway. In this « Corporate Memory » project we were in charge of developing a « Personal Memory Assistant », (2) the OCTO project , for the French Ministry of Defence, whose acronym meant « Conceptual Objects of the Battlefield ». Its aim was to develop a tool to help headquarters officers represent complex situations and take decisions.

By 1998, we started to market a personal version of Ideliance to large companies, like L'Oreal, Air Liquide, France Telecom, and to the French Army. Following pressure from customers, we developed a client-server version of Ideliance, allowing small communities of persons to share semantic networks.

In 2002, lacking financial resources for its development, the Ideliance company had to stop, and the product was underpinned by Thales, a very large Defence and Electronics company, where it is mainly used for Security and Military Intelligence applications.

Main contributors of Ideliance were Sylvie Le Bars (www.arkandis.com) and myself for the design, and Stéphane Jean and Denis Poisson for the implementation.

There are few available academic references about Ideliance, because it was first developed in a SME, then its main applications were in the Defence domain. Some references are given at the end of this paper.

In some sense, Ideliance was a fruit of the AI winter, and made a bridge between the AI times and the current Semantic Web era. A deep knowledge in AI, Logic Programming, Knowledge Engineering is useful for the design of future desktop systems.

MAIN DESIGN PRINCIPLES

Principle 1) : Simplicity. The main idea is to let users write semantic networks for themselves, to express any information of interest in their activity. The tool must be simple enough to allow any person capable of structuring its thoughts to translate it into a semantic network. We call « collection » a given semantic network, following the rule that concepts and vocabulary to describe and use Ideliance should *absolutely not* refer to technical or theoretical terms of computer science. With the Ideliance editor, users can create *Subjects*, assign them to *Categories*, and link them through *Relations* to constitute *Statements* like « *Jim / works for / W3C* ». All relations have a reverse relation, so that the reverse Statement « *W3C / employs / Jim* » is built automatically. Subjects in statements can also be dates, files, URLs or mail addresses. Text notes can be attached to a subject. There are no type or cardinality constraints.

N.B. An intermediate name for Ideliance was *Idécriture*.

Principle 2) : Emergence. There is no a priori declaration of structures, models, templates, classes ... All these notions emerge automatically from the reality of the semantic network : if you want to describe a new Person, Ideliance makes you suggestions computed from the current set of information : relations most frequently used to describe a person, and most frequently used subjects used to complement them are proposed as a first choice to the user. This principle of emergence, learning, discovery is used thoroughly in all aspects of the user interface.

Principle 3) : Powerful Query and Reporting tools : There is in Ideliance a plethora –although of various levels of interest and usage- of tools to ask queries, draw relations graphs, build dynamic OLAP reporting arrays, from the semantic network, and to publish it in various document formats.

Principle 4) : Gateways with usual data formats : Structured information from tables of spreadsheets and relational databases is easily translated into the semantic network format, and the other way round.

Principle of (Simplicity + Emergence) is a nice compromise. It lets novice user start with simple things, without the burden of a-priori modelling. It allows also to start with simple Semantic Desktop engines, which can be later improved by adding more sophisticated learning and discovery algorithms, in a smooth transition mode for the user.

Gateways with other formats is key to have users accept putting critical information in the Semantic Desktop. And Query and Reporting tools should compare with what they can get from tools like Access, Excel, Crystal Report. We experienced repeatedly that the main competitor for a Semantic Desktops is ... MS Excel ! We will address the role of a Semantic Desktop vis à vis the Semantic Web in another point.

A key –yet difficult- point for the success of Semantic Desktops is to persuade users that it isn't more demanding in terms of effort and intelligence than using Excel, and that the benefits will increase tenfold.

APPLICATIONS / MARKET SEGMENTS

We found the following types of applications, which can be considered as market segments for Semantic Desktops in general :

--general purpose permanent personal information management

Individuals decide by themselves to store and retrieve information strictly under the form of semantic networks. For instance, a business developer maintains a track of all his contacts, meetings, bids, competitors. He currently has a collection of more than ten thousand subjects, split into a dozen categories. He also maintains collections about his roadrunner activity, and on building his new home.

--specialised temporary personal information management

Consultants have used Ideliance to perform missions (like assistance to tenders, audit) and to deliver their results to their customers. Instead of receiving classical static Word + Excel + Powerpoint documents, their customers can dynamically browse, query and draw graphs of the current state of the consultant work.

--knowledge management in small groups

Merck Europe Pharmaceutical Labs have used Ideliance to formalize and develop a multidomain encyclopaedia on their technical, research, marketing and industrial property. They designed themselves the underlying ontology, using the semantic editor. It really changed their way of interworking among departments.

--online critical information sharing

L'Oreal, Air Liquide, Thales used Ideliance to formalize and collect information on competition. The design, formalization and collection of such information is mainly an individual task, while consultation is collective. This is a dramatic change as compared to trimestrial static documents about competition.

-- publishing the contents of classical databases in Semantic Networks

At l'Oreal, classical patents databases are translated nightly into semantic networks. The Semantic Networks facilities for browsing and querying offer, at zero development cost, a much more sophisticated interface than the original database.

-- Intelligence Applications

Ideliance is currently in operational use by French Army in France and abroad.

Since 2004, we chair the the NATO IST-38 Research Task Group on Information Fusion Demonstration, where Ideliance is one of the instrumental tools. This group is conducting an extensive study of the interest of semantic representation for Intelligence Analysts.

All these applications share the following property : users are pleased to develop critical, high value information systems themselves without the cost and lack of flexibility of traditional IT applications.

USERS ACCEPTANCE

One of our first discoveries was that users were very sensitive to the vocabulary brought by the tool. In the first versions, inspired by computer science and philosophy terminology, we used terms like « entities » « characteristics » « models ». This was strongly rejected by the targeted customers (chemist engineers, biologists, officers, medicine doctors, sales persons ...). We made many evolutions, and the current vocabulary uses : Subject, Relation, Category, Statement, Collection. Yet we may have to go some steps further, down to Subject, Verb, Object, Sentence Needless to say, the term « ontology » is a repellent !

In the Business and Industry world, « going semantic » would be a revolution. People use the web, email and MS Office tools, and they are lightyears from the idea of leaving their daily universe of documents. The good news is that there is a small proportion of individuals who immediately adhere to this « new way of writing ». Such people, as individual or small groups, can switch very rapidly -within days- to it, and they very seldom abandon it. Any marketing policy should leverage on these early adopters (and often true addicts !).

But the huge majority of people first rejects this new approach. To tackle the challenge of users acceptance, we propose to consider two « extremist » and opposed strategies : « Extreme Explicit Semantics » and « Extreme Implicit Semantics ».

The « *Extreme Explicit Semantics* » puts forward the idea that people should write information *directly in semantic networks*, as a substitute to producing Word, Powerpoint, Excel documents, even emails

The advantage of this radical approach is to set very high constraints and challenges for the design of extremely user-friendly editors. A strong consequence is also that this new kind of writing should probably be considered as a new cultural asset. The usual idea that « computerized tools should need no training, be understood and used within 5 minutes, unless they are rejected » is here completely out of the scope. People spend months and years to learn arithmetics, grammar, foreign languages, so will be the case for the semantic writing.

We experienced that a minimal course for Ideliance users lasts two days. And an ideal semantic project building session for a group lasts one full week. (of course Masters students need only 15 minutes to get fully acquainted with the tool, but have no idea or motivation for its usage).

We also felt the need to develop a methodology to teach people what semantic networks are, with a one day session *without computers* but with post-it, blackboard, games of language, and hand-written exercises. We experienced that the time lost in this day was rapidly saved later.

Indeed, in this domain, a special attention must be given to the relationship between Semantic Desktops and all the accumulated folklore, experience and tools in the domain of MindMapping.

The «*Extreme Implicit Semantics*» approach takes the opposite direction : we leave users to their favourite Office tools, and smart programs analyse the content of these documents, understand them, and implicitly, automatically build the corresponding semantic network ... Whereas the *Explicit Utopia* puts all the effort on the user, the *Implicit Utopia* challenges computer scientists with the Natural Language Understanding problem. With more realism, this challenge is today approached by some text mining and automatic tagging tools.

Not surprisingly, we consider that, for future Semantic Desktops, these two approaches, Explicit and Implicit, should be accommodated. We have good reasons to hope that they will cross-fertilize : text understanding will be much easier in presence of pertinent contextual semantic resources, and the production of semantic networks will be quite accelerated by linguistic tools capable of finding or suggesting subjects and relations from texts.

Inside Thales, we have started to experiment several prototypes of bidirectional tight or loose connexion between text mining and semantic networks.

We consider that this cooperation of explicit and implicit semantic approaches is a key technology for the future Semantic Desktops.

Note : during the years of development of Ideliance, we implemented a lot of advanced built-in features (logic programming, constraints programming, sophisticated graphs analysis algorithms). Although these features were very exciting to develop, we experienced that it was difficult for normal users to adopt them.

The first challenge Semantic Desktops will have to face is to make « semantic writing and reading » popular, before starting to complexify it. And the first step is education. It is not a « Click and Forget » application.

IMPLEMENTATION

Since 1993, we had the opportunity to experiment several alternatives for the implementation of a Semantic Network Manager. The corresponding engine has to support strong constraints : the structure of information is more dynamic than for relational databases, algorithms –on graphs for instance- are more demanding than just sorting, joins and set theory ones. Other features like renaming subjects, fusion of collections, implementation of emergence, learning and discovery features, are uncommon.

From the lessons we learnt, our recommendations for the implementation of Semantic Desktops are the following :

- put data very close to the screen of the user - in the main memory of the desktop - and design the implementation accordingly. If you also need a light client / server implementation, upload the heavy client design to the server rather than the opposite
- consider that in the future your design will have to accommodate semantic networks filling hundreds of gigabytes of fast, random access solid state memory
- low level data structures should be as simple –even naïve- and regular as possible.
- consider persistency (on disk) as a secondary problem, to be solved separately
- the physical information representation should be fine tuned, and mastered down to the bit level. Do not trust others –like Java or Relational Databases- to take care of the rock-bottom performance.
- said differently : Semantic Networks are at the same level as both Object Oriented Languages and Relational Databases. In the same way Java is not implemented above Java and Oracle not implemented above Oracle, do not implement Semantic Networks –if you want to handle future large networks in real life - above Java and Oracle. In the case you should anyway use such high level tools, restrain yourself to their simplest features.
- the physics of XML-based syntactical formats (or RDF, OWL, ...) should not be taken into consideration for internal implementation. Their interest is in standard external exchange formats.

Semantic Desktop Computing must be considered as a very innovative, -some may say revolutionary- way to build and use information systems. In this respect, it demands and deserves implementation techniques of its own.

ROLE VISA VIS A VIS THE SEMANTIC WEB

One can play many games with the three words : SEMANTIC / WEB / DESKTOP, and « Semantic Desktop » is just one of the games. We can also consider the following ones with interest:

(SEMANTIC WEB) DESKTOP
SEMANTIC (WEB DESKTOP)
(SEMANTIC DESKTOPS) WEB
etc ...

It strikes us that some leading Semantic Desktop projects like Haystack, Gnowsis and IRIS also find their motivation in personal information management, before looking at the global (semantic) web resources, although they borrow from it the basic semantic network (« triple ») representation scheme. In this respect, we can say that the Semantic Desktops future could be absolutely independant from the existence, strength and prosperity of the Semantic Web. (After all, semantic nets –as for them- were invented by Aristotles). Ideliance illustrates this point of view, since it started even before the Web had come to a visible existence.

It is also striking to realize that the motivation for the Semantic Web was to let machines exchange information globally, whereas the motivations for the Semantic Desktop is to let humans organize information locally!

Should we consider that a P2P protocol between Semantic Desktops is still part of the Semantic Web ?

However the synergies between Semantic Personal Tools and the Semantic Web are numerous :

- a Semantic Desktop should include a pure « Semantic Web Browser » : as « read only » as Web Browsers are
- all the ongoing work on sharing, aligning, updating ontologies on the Semantic Web may provide useful guidelines at the smaller scale of the Semantic Desktop
- of course, semantic nets produced by Semantic Desktops will feed the Semantic Web –provided that efficient policies to map them with Web Ontologies are designed
- on the other hand, Ontologies and RDF producers in the Semantic Web will be motivated by all these new « clients », and also care more about the quality and usability of their data

CONCLUSION : TOWARDS « Intelligence Amplifiers »

There are many other aspects in our experience with Ideliance which we could consider in the light of the Semantic Desktop paradigm : notions of collections fusion and extraction, labelling triples with access rights, embedding text search facilities ...

The main lesson we would like to transmit is that *Semantic Desktop Computing is about people, more than about machines, architecture and protocols*. Users acceptance is key, and we badly need projects which, with the help of human and social sciences, study the cooperation between humans and Semantic Desktops.

At the end, Semantic Desktops will become « Intelligence Amplifiers »

REFERENCES ABOUT IDELIANCE

Volle Michel. Idéliance, ou "comment se faire aider par l'ordinateur pour réfléchir", in

<http://www.volle.com/opinion/ideliance.htm>

(A MIS consultant reports on his discovery and usage of Ideliance as a Personal Information Management Tool)

Rohmer Jean. Représentation, Fusion et Analyse d'Informations mises sous forme de Réseaux Sémantiques : vers le « Calcul Littéraire", in « *Les systèmes d'information élaborée* » Société Française de Bibliométrie Appliquée, Ile Rousse, France 14-18 Octobre 2002

Rohmer Jean. The Case for Using Semantic Nets as a Convergence Format for Symbolic Information Fusion in NATO RTO-MP-IST-040 Information Systems Technology Panel (IST) symposium on « *Military Data and Information Fusion* », Prague, Czech Republic, 20-22 Oct 2003.

Bruneau Jean-Maurice. Ideliance, Logiciel de Rupture pour l'Intelligence Economique? Un Cas d'Application sur les Signaux Faibles, in VSST(*Veille Scientifique Stratégique et Economique*) 2001 Conference Barcelona, Spain