

Wiki and Semantics: Panacea, Contradiction in Terms, Pressure for Innovation?

Some experiments and tracks towards *Intelligence Amplifiers*

Jean Rohmer

Centre des Nouvelles Technologies d'Analyse de l'Information
Thales Communications, France
jean.rohmer@fr.thalesgroup.com

Abstract. This paper examines the relative characteristics of wiki principles and of semantic systems. It first stresses some oppositions between these two approaches, and exposes the challenge of their reconciliation. We then make a detailed description of Ideliance, a „pure“ semantic tool, and we set criteria to compare several existing semantic wiki systems. After a critical look at some of their features, we propose precise directions for cross-fertilisation of semantics and wikis, advocating for solid, long-term foundations.

1 Semantic Wiki: an oxymoron which raises many questions

We must first realise that the existence of the „Semantic Wiki“ concept, comes — like many other Information Systems concepts — from our inability to build machines or programs which automatically understand natural language, either in the form of documents or in the form of spoken or written conversations (as quoted in [1], best analysis programs fail to understand a sentence in more than 70% of the cases). Machines cannot help us without „semantics inside“, and today we have to strenuously feed them with this semantics.

In practice, we permanently have to waver between textual document management and structured database applications. *Is there a life between Word and SQL ?* That is the question Semantic Wiki is about.

There are two ways to approach this question. The first one is extremely theoretic, this is *semantics*, the second one is extremely practical, this is *Wiki*. At first glance, they seem to be much too distant to hope that any fusion is possible. Semantics has been endlessly studied for millenniums in literature, philosophy, philology, linguistics, and Wikis sometimes look like an odd tinkering from an idle programmer's week-end. Moreover, while Wiki is the Hawaiian word for „quick“, semantics refers to things like syntax, study, grammar, school ... and school comes from „skole“ which is the Greek word for „being slow“, „not to hurry“ !

In other words:

Semantic Wiki means Slow Quick (SlowQuick ?)

In this paper, we would like to address various facets of this contradiction, at the light of our experience in the development of a Semantic Network Manager: IDELIANCE [2] and of Information Technology usage in large and small Business environments.

In fact, Semantic Wiki is an ambitious endeavour: it aims at increasing the synergy between people intelligence and the power of the grid of networked computers; and in this period, where it is still embryonic, we should take time *-skole-* to be sure to start with the right groundwork.

In this paper, we first describe the main features of IDELIANCE at the Wiki light, then we review the main characteristics of significant Wiki implementations. This will help us to reformulate the Semantic Wiki challenge, and to make proposals for sound directions for the future.

2 Ideliance: a „pure“ semantic network manager with some wiki properties

As reported in [2], Ideliance was initially an attempt to bring the best ideas of Artificial Intelligence on everybody's desktop. In the same way as Ward Cunningham [3] wanted to develop the „simplest online database“, we wanted to develop the „simplest personal knowledge base“. For that purpose, we chose to develop a user-friendly Semantic Networks [4] editor for everyday use in professional life. The first Ideliance prototype was available in November 1993. Commercial usage started in 1996 for France Telecom. Off-the-shelf personal versions were bought in 1998 by French Atomic Energy Commission (CEA), and French Army. The server version was installed in 2000 in large corporations like L'Oreal and Merck Pharmaceuticals. [2] reports on lessons learnt from Ideliance applications.

Although the question of finding the right balance between textual and structured representation was perpetual during Ideliance design, the initial choice was clearly in favour of *full structured semantic networks*. Ideliance proposes users to create **subjects**, belonging to zero, one or more **categories**, and to write **statements** of the form (subject / **relation** / subject), where each relation has an **inverse**. A complement in a statement can be not only another subject, but any resource (file, email, URL). A subject is any character string (including numbers and spaces), without comma, and without length limitation. A set of Ideliance objects is called a **collection**. Example of a statement and its inverse:

John Paul Wagner **works for** United Nations (UN)

United Nations (UN) **is the employer of** John Paul Wagner

In fact, and this is a main difference with wikis, *there is no textual format for statements*: all statements are built from a graphical interface which let users

manipulate only names of subjects, relations and categories in a controlled way. Nevertheless, we authorise a free text area associated with each subject, with zones which may point to other subjects, without label/ relation on this link.

The basic mode of information display is the Ideliance „Card“: given a subject, it collects and displays all the statements having it in subject position. This allows for an immediate navigation mode from Card to Card, each being built dynamically. In the server mode, a card displays all up-to-date statements about a subject written by any user. More precisely, each statement has a signature (author, date of creation, visibility rights). User see only statements they are allowed to, either directly or through group membership.

Conversely, several tools allow to publish Ideliance contents (e.g. cards) in useful standard formats like Word, HTML, XML, Excel. Utilities permit also to convert XML and Excel files into Ideliance statements. All reasonable features we can expect on objects are available: delete, rename, duplicate, merge, extract. For instance we can extract and merge collections.

A difference with most of Semantic Wikis is that there is no reference to a standard like RDF in its implementation. Not only because it simply did not exist at this time, but because we wanted *to make symbolic statements the unique, atomic concept for information representation*, totally in the hands of end-users, so as to close all the backdoors to Software Engineers for any underground traffic on information. This does not impede the usage of RDF as a *standard interoperability format* between Ideliance and other semantic applications. (There is no mandate to choose RDF as an *internal* implementation feature). A WSDL Web Service has been developed to let other applications read, write, query an Ideliance server. These services could be reused to provide a more Semantic Web compliant interface to Ideliance.

The main **radical** idea of Ideliance is to **try to** get rid of the notion of document. Each atom of information is a statement, and the tool collects and displays them on demand.

The card feature is one of these tools, but many others are available:

- semantic queries
- textual queries (à la Google)
- simple or complex tables in OLAP-like style
- simple or complex graphs

None of these features has a textual format accessible by the users, who can only go through an interface. (Internally, the objects corresponding to queries, tables, graphs are represented as system-visible Ideliance statements).

Textual queries simulate the notion of document through the dynamic cards. They retrieve all subjects whose card content matches the query. An option is to extend this textual search to the content of documents used as complements to the card subject. For this purpose, Ideliance embeds the Wilbur search engine ¹.

Ideliance interaction with the user fully relies on the notions of **emergence and suggestions**, which can be generally stated as: when an user starts an

¹ see <http://wilbur.redtree.com>

action, the system suggests the usual ways to continue / complete it. We will develop this point later in the discussion.

All these specifications and implementations were not done overnight. Ideliance is the result of many years of evolutions with many kinds of users, either individual or collective ones.

Currently, we are experimenting the addition of information analysis tools: data mining, clustering, knowledge discovery, rules, in particular for Military Intelligence and Business Intelligence applications.

3 A survey of some current Semantic Wiki proposals

Existing semantic wikis can be compared using some alternative options:

Approaches to the challenge of accommodating both free text and structured semantics:

- put structure / formalism inside text (option A1)
- put text inside structure / formalism (option A2)
- exclude text (option A3)

Note that the second approach is the one adopted by HTML, as the essence of the Web, and later by XML.

Global design approach:

- grounded in technical architecture choices (option B1)
- driven by an end-user perspective (option B2)

As we have seen before, Ideliance illustrates Options A3 and B2.

Platypus [5], is a good illustration of option B1 : the idea is that a Wiki page is annotated -in hidden fields- by RDF metadata statements, and that an HTTP server can selectively download these metadata to the client, allowing a natural chain of navigation. In Ideliance, we implemented a similar feature „Ideliance Inside“ for HTML pages, where the hidden (and proprietary at that time) Ideliance format could be extracted from an HTML page, along with a symmetric mechanism allowing to generate Ideliance cards in human-readable HTML, accompanied with the embedded corresponding Ideliance collection.

WikiSAR [6] is an illustration of option A1: the subject is implicitly the page name, and, on a text line, verb and complement are indicated, separated by a colon, each of them following the WikiWord conventions. This very simple structuring scheme is complemented by the capability to insert formal queries *forged with the verbs* in the text. They then are replaced by their result *always up-to-date*. Moreover, WikiSAR can visualise the network of sentences through a friendly graphical interface.

Rhizome [7] is clearly a technical architecture (option B1), with the idea to implement a sort of algebra on subsets of RDF triples. It introduces specific

formal languages -ZML, RxML, which make Rhizome a flexible workbench to manipulate semantic nets.

Ikewiki [8] outlines general, ambitious goals for a collaborative semantic environment. Authors analyse in a systemic way the relationship between users acceptance, expressiveness, generality of semantic applications, following a B2 option. More than a Wiki, this project is a general purpose, users-oriented platform for semantic information processing.

Semperwiki [9] is a simple, personal semantic system, with an option A1 structure-in-text approach similar to WikiSAR, and a strong emphasis on bringing incentive to the semantic effort of the users (a B2 option).

A broader discussion of the Semantic Wiki field should also encompass the relationship with the *very close* Semantic Desktop domain [10].

4 Some surprising things about Wikis and Semantic Wikis

As stated in a previous paragraph, the Semantic Wiki is today a tiny domain, but, in the same time, it sets some key challenges for the evolution of Information Processing systems. It is de-facto a domain which harnesses all studies conducted since more than 40 years in Software Engineering, Artificial Intelligence, Formal Logic, Natural Language Processing. We must not take a narrow, short term view of it. Instead, we must analyse this domain lucidly.

The tinkering aspect of some Wiki implementations (see for instance *TheManySetsOfRulesToBuildWikiWordsAlsoCalledCamelCase*) could at the end of the day cause more harm than benefit, even if they contribute to the very nature and initial success of wikis. A long term perspective for Semantic Wikis needs a joint, sustainable effort from the community. At this point the *semantic side* of Semantic Wiki should help: it is striking to realise that explaining the basics of a semantic tool like Ideliance finally takes less time and space than detailing all the folklore around Wikis. (WikiTag, WikiCategories, QuickSurvey, ReverseIndex, RoadMap, WikiSingleWordProblem, WikiNamePluralProblem, WikiKeyWords, etc, etc). Finally the written documentation on the Wiki unformal conventions is thicker than the formal one on semantic tools. *In this respect, we think it is safer to carefully inject the Wiki spirit inside Semantic (Desktop) systems than the opposite.*

A surprising thing about Semantic Wikis is the quasi absence of reference to the tools used daily by all professionals: Excel, PowerPoint (or their open source equivalent). Such tool are in themselves *excellent structuring tools which embed a lot of semantics* as compared to textual documents. We should make efforts to bridge their *actual* semantics with the *potential* semantics of wikis, all the more if we target personal or collaborative applications. A side point would be the support of figures and simple arithmetic, which are ubiquitous in any kind of business. If we do not address these points, the risk is to limit ourselves to the development of encyclopaedia-like usages.

I am always very surprised to notice that the notion of *inverse relation* is nearly absent in semantic tools. It was one of our first decisions in the design of Ideliance. This trivial mechanism has nevertheless the capability to build automatically symmetric cross-referencing and navigation from one card / page to the other. It is also extremely useful to express queries and rules in a symmetric way.

Finally, the constant reference as a kind of creed to RDF in most Semantic Wiki designs seems to me overrated: A Semantic Wiki is *neither more semantic nor more wiki* simply because it gives visibility to the RDF standard. A Semantic Wiki user in the future should ignore the existence of RDF, as well as PowerPoint users ignore that there exists a Metafile format which allows them to cut / paste schemas from / to other office documents formats. *RDF is key to provide interoperability among the semantic applications, but is an irrelevant concept for the end-user.* For instance, the rock-bottom notion of blank node is a semantic nonsense for the end-user. In the same spirit, the notion of unique URI, often quoted as an advantage when placed inside a RDF statement, is a concept which becomes less important for users, more familiar with information retrieval through search engines than through the URI of their object of interest. *The semantics as perceived by users should differ from the one perceived by programs.* An argument for this statement is that many wiki designers forge new friendly formal dialects, softer than RDF, to let users manipulate formal semantics. (See for instance ZML [7])

5 Proposals for a sustainable fusion of Semantic and Wikis

In the previous paragraph, we were voluntarily critical at some aspects of wikis, semantic wikis and semantic systems. We would like now to make proposal to cross-fertilise their advantages while minimising their potential weaknesses.

Transpose to semantics the wiki culture of a community taking care of a knowledge-base The success of Wikipedia is the proof that some users are diligent towards informal semantics. Let us encourage the same sort of people to become as diligent towards formalised semantics. There exist *potential semantic (re)writers / translators* which could translate natural language pages of a wiki into a semantic format (in the same way publishers are used to have papers translated from one natural language to another). Such added value would generate a networked snowball effect : reading the wiki would become more pleasant and efficient, and casual writers would be tempted to turn themselves into *semantic writers* to raise the level of retrievability and understandability of their contribution.

Design carefully a specific level of semantic formalism for the end-user As exposed above, this level should definitely depart from the Semantic

Web standards (RDF, OWL) which, by essence, were semantic constructs for machines, and not for people. Of course, this user-level semantics could be made operational and interoperable through the use of the Semantic Web standards, but in a hidden way, in lower layers of the Internet machinery. This level should also differ from the first generation of wiki conventions, while keeping their freshness.

Our experiment with the simple Ideliance formalism is a proof of existence of this level. Designing and agreeing upon such a user-level semantic will not be immediate, but it is a long term key success factor.

Make Semantic Wikis a companion of office tools instead of a substitute This is mandatory for the acceptance of Semantic Wikis in most of economic sectors. Not only Semantic Wikis should import / export their contents from / to office documents formats, but they also should be able to capture in real time the semantics of graphical editing a presentation or a spreadsheet. In the same way current wikis have a Text Processor-like face, there must exist Semantic Wikis with a Spreadsheet-like face, a Presentation Editor face.

With the emerging XML standards to describe the contents of office documents, this objective is not out of reach.

Encourage a „Semantic Inside“ policy for HTML pages The idea of populating HTML pages with semantic statements, along with the capacity of Semantic Wikis or Semantic Desktops to selectively download them seems very simple, and capable of initiating a viral propagation of semantic statements. It has been implemented as an „Ideliance Inside“ feature, and also in [5]. We should analyse why it does not exist in reality, and what conditions should be met to start such a proliferation.

Use human-driven discovery and emergence mechanisms for vocabulary / ontology congruence The whole semantic game is complex : it is a continuum of interactions between a continuum of levels, e.g. from personal, to workgroups, to corporate, to global level. The notion of ontology, ubiquitous in the Semantic Web along the idea of URI uniqueness, needs to be reformulated to meet this complexity.

„But where danger is, grows the saving power also.“ cited from „Patmos“, by Friedrich Holderlin.

Here the „danger“ is „people complexity“, as compared to the simple „machines“ world of the Semantic Web. And thus the saving power is „people“ too.

The alignment / congruence tools in the semantic wiki world should *empower users with the ultimate decisions concerning the meaning of terms*, instead of being blind black boxes trying to make some „optimal“, global ranking of the „best“ meaning. These tools must compute / discover emerging properties from the whole knowledge base, let users make their choice, observe these choices, and capitalise from them for further recommendations.

In Ideliance, a first modest implementation of this emergence principle consists of maintaining statistics on relations and complements of subjects of a given category. This simple mechanism has already the capability to give, in real time, an up-to-date view of the „data model“ of the collection, without any a-priori declaration.

In the same way the Sun Microsystems slogan is „*The Network is the Computer*“, we tend to say: „*Information is The System*“. The system structure emerges from the information it contains.

Take into account higher levels of semantics: Discourses, Argumentation, Rhetoric Any *semantic* writing act has an intention, an objective, expects some results or benefits. Personal or collective semantic tools should help users in such high-level tasks. This is for instance by the ABCDE format for scientific publishing [1]. This paper is a first attempt to cleave the monolithic document, by explicitly labelling paragraphs as **B**ackground, **C**ontribution and **D**iscussion, along with general **A**nnotation, and collection of useful resources in an **E**ntities paragraph.

The other nice side of this *editing* effort will be the *reading* side, which, after some *thinking*, and *working*, will again yield a new *editing* activity.

We could call this proposal „*soft semantics*“, since it keeps the very meaning inside natural language sentences or paragraphs. It makes a step further than *metadata semantics* à la Dublin Core, which takes the document as a monolith. With Ideliance, we clearly are trying a „*hard semantics*“ approach, which we call also *Extreme Explicit Semantics*. (The *Extreme Implicit Semantics* would be automatic natural language understanding)

If we merge „*hard semantics*“ with the ABCDE approach, we could contemplate tools which not only would represent the semantics of each sentence, but also the semantics of the *making-up of all the sentences* towards the author's intention. This would permit to represent things like: „statement A is used by author B as an example of statement C which is later used as a support for statement D“. In a companion paper [11], we outline the notion of „*litteratus calculus*“ as an infrastructure to underpin such an approach.

6 Conclusion: Towards Intelligence Amplifiers

[6] advocates for immediate gratification for semantic writing. This idea of a fast ROSI (Return On Semantic Investment) is well in pace with the Wiki world. And this point is key for the acceptance of semantic wikis.

Coming back to our built-in contradiction (Semantic Wiki = Slow Quick), we would like to point out another kind of gratification distilled by semantic writing: it is a long term, deferred, -*skole*- reward : the pleasure of thinking, of installing order, clarity in one's knowledge and thoughts. The effort to decide to create a new subject, to choose a category for it, to forge a new statement linking two subjects is a mental exercise, which, day after day, amplifies *your* intelligence. Clicking is not thinking.

Acknowledgments:

I would like to thank the main contributors to Ideliance : Sylvie Le Bars² for designing the subtle tempo between slow and quick, click and think, and Stéphane Jean and Denis Poisson for the technical design and implementation.

References

1. Waard, A.D., Oren, E., Haller, H., Völkel, M., Mikka, P., Schwabe, D.: The abcdef format. In: Submitted to ESWC 2006. (2006)
2. Rohmer, J.: Lessons for the future of semantic desktops learnt from 10 years of experience with the ideliance semantic networks manager. In: ISWC 2005 Galway, Ireland. (2005)
3. Cunningham, W.: (Wiki design principles) available at <http://c2.com/cgi/wiki?DesignPrinciples>.
4. Sowa, J.F.: (Semantic networks) available at <http://www.jfsowa.com/pubs/semnet.htm>.
5. Campanini, S., Castagne, P., Tazzoli, R.: Platypus wiki. In: 3rd International Semantic Web Conference (ISWC) Hiroshima, Japan. (2004)
6. Aumueller, D., Auer, S.: Towards a semantic wiki experience – desktop integration and interactivity in wikisar. In: ISWC2005 Semantic Desktop Workshop, Galway, Ireland. (2005)
7. Adam, S.: Building a semantic wiki. *IEEE Intelligent Systems* **20**(5) (2005)
8. Westenthaler, R., Schaffert, S., Gruber, A.: A semantic wiki for collaborative knowledge formation. (In: Semantics 2005, Vienna, Austria, 24th November 2005)
9. Oren, E.: Semperwiki: a semantic personal wiki. In: Proceedings of the 1st Workshop on The Semantic Desktop, Galway, Ireland. (2005)
10. First Semantic Desktop Workshop 2005, Galway, Ireland, S. Decker and J. Park and D. Quan and L. Sauermann (2005)
11. Rohmer, J.: Litteratus calculus: a manifesto for a demographic way to build a sustainable semantic web. In: ESWC. (2006)

² see www.arkandis.com