SPINE – Semantics beyond search

Bill Howell, versions sent out: 1st draft based on material from home 19May11, 10Aug11 (exercise only), 05Sep2011 this "almost complete" draft was posted on www.gcpedia.ca, 06Oct11 – copyright & waiver improved

Summary

While "Semantics" is the concept central to this paper, the overall objective of this paper is not to provide a review or roadmap for semantics experts who are building toolsets and systems, but rather to greatly expand the awareness of the social media and collaborative communities concerning the "larger world of capabilities" beyond improved search engines that semantics will help to enable. Furthermore, it is my gut feel that there are fundamental limitations to semantics (and logic), so it is important to be aware of these, and to be aware of completely different concepts and toolsets that pick up where semantics and logic fall apart.

With respect to building social media capabilities into government IT systems, given their very immature state at present, rather than focus on specific ideas it is more important to consider **3 themes**:

1. Identify & cull a wide diversity of concepts that are being proposed or implemented around the world; and given this understanding…
2. Plan and develop IT systems to allow for the easy integration of new toolsets, processes and systems, many of which will be completely unknown and unpredictable at the time of integrating current capabilities (such as Facebook, Twitter, etc) into government systems. In this light, it is more important to design our IT systems to be open to accommodate new unexpected capabilities, rather than to provide a shopping list of semantic or social graph capabilities to select from.
3. Allow for "multiple conflicting developments" – demonstrations to try out many ideas and see what works in the "social media marketplace". It for to easy to over-emphasize analysis and planning when both are of limited effectiveness.

This can be considered as the main recommendation from this report. Although systems design issues are discussed in a separate paper, hopefully it will be clear from this paper that the utilization of new semantic capabilities will be helped and/or hindered by the systems mindsets (conventional or open).

It is my assumption that the most important social media capabilities will be neither predictable nor well understood in their initial stages. Within that context, the current document is a limited, random look at several potential tie-ins between semantic technology concepts and work-environment applications. A second document will deal with social graphs and social sets ("… the heart of Facebook …"). **In both cases the intent is to push the three themes above.** A third document is planned for criteria and constraints of IT systems to achieve openness, flexibility, reliability, and robustness.

I recommend that you simply look at this Summary, Part I, but ignore the rest of the paper (Part II) unless it really is of interest to you. In other words, it's a bit like walking a tightrope: "Don't look down!" For shear fun, you might want to test friends on the "Plausible next sentence exercise" (Appendix D gives a summary, but the exercise appears in a separate document).

Note that "Semantics" and the concepts in this paper are "established" technologies, albeit under rapid development, as compared to the much more recent "social sets and social graphs" technologies, especially when one considers the dynamics of the latter. Furthermore, it is the "social sets and social graphs" that are at the heart of the new social media. They are the intended focus of the next paper, which is a much more difficult challenge than the semantic one.

Endsection

****

****

Waiver/ Disclaimer

The contents of this document do NOT reflect the policies, priorities, directions, or opinions of the federal government of Canada, nor even of the author and any of his past current, or future employers, work colleagues family, friends, or acquaintances. The contents (including but not restricted to concepts, results, recommendations) have NOT been approved nor sanctioned at any level by any person or organization.

The reader is warned that there is no warranty or guarantee as to the accuracy of the information herein, nor can the [analysis, conclusions, and recommendations] be assumed to be correct. The application of any concepts or results herein could quite possibly result in severe losses, damages, and death to the readers, their associates, organizations, municipalities, states, countries, civilizations, religions, and the entire human species.

The author and the federal government of Canada accept no responsibility for damages or loss arising from the application of any of the concepts herein, neither for the reader nor third parties.

endpage

Table of Contents

****

Status as of 10Oct2011:

- Several concepts/ issues are very incomplete – notably data mining, Latent semantic analysis, and other concepts related to semantics.
- I will not provide details for the "Connectionist" concepts, as my time will be better spent working on the "Social graphs and social nets" paper. However, the listing is an important placeholder for future elaboration, as many of the concepts are of direct relevance.
- Frankly, there are a huge number of relevant concepts and I can't cover them all in a short paper. However, as the intent was to "push the thinking" on Semantics, what I have included should have that effect for individuals having the right basis and interests.

****

CAUTION:

Please note that this document is largely drawn from home projects and posting of Bill Howell. It is adapted here to the context of social media as per the SPINE project at NRCan. There is therefore a strong personal bias and background perspective to this document, which limit its generality, and I am several years out of date on this subject area. The content is also NOT NRCan property, as components are from Springer-Verlag, and other components have been posted on my website, and pre-date SPINE by several years (in particular the section on Confabulation theory and the survey that I did in 2006-2007 based on material from 2002 and 2007).

DOUBLE CAUTION: Semantics is NOT my field, nor have I made a special effort to read into it extensively. My own exposure has come through the papers and discussions of others, primarily via computational intelligence conferences, journal papers, and emails.

endpage

Part I - Semantics, Social Media, and Workplace Applications

Introduction

"Semantics" (see the section "Definitions of Semantics") has finally become a fashionable word in collaborative systems and social media systems. But its usage has largely been restricted in popular articles to providing improved search capabilities "à la Google" and beyond.

While "Semantics" is the concept central to this paper, the overall objective of this paper is not to provide a review or roadmap for semantics experts who are building toolsets and systems, but rather ***to greatly expand the awareness of the social media and collaborative communities concerning the "larger world of capabilities" beyond improved search engines that semantics is helping to enable***. Furthermore, it is my gut feel that there are fundamental limitations to semantics (and logic), so it is important to be aware of these, and to be aware of completely different concepts and toolsets that pick up where semantics and logic fall apart.

It is my assumption that the most important social media capabilities will be neither predictable nor well understood in their initial stages. This isn't such a brave statement given that our use of current technologies is hamstrung by not fully recognizing important long-standing capabilities and constraints. For example, end-user definable access control and text search capabilities are still poorly appreciated by IT specialists and end users in spite of several decades of examples, and several decades of "crippled applications". Think of how often "collaboration" is limited to a barrage of emails back and forth to members of a community or project, then think of how messy it is to track and check that information when you don't have the time to file, summarize and arrange it.

Concept development is best described as a evolutionary process in an open competitive market, with most of the key breakthrough ideas coming from non-institutional sources (eg some [hacker, student, academic] on a computer at [home, school/ university, company]). Within that context, the current document is a limited, random look at several potential tie-ins between semantic technology concepts and work-environment applications (see the section "Semantic applications to the workplace – random thoughts"). A thorough literature review has NOT been done, and most of the thinking is based on my own hobby interests at home, combined with news items that have arisen within the "SPINE" project at work.

With respect to building social media capabilities into government IT systems, given their very immature state at present, rather than focus on specific ideas it is more important to consider **3 themes**:

1. Identify & cull a wide diversity of concepts that are being proposed or implemented around the world; and given this understanding…
2. Plan and develop IT systems to allow for the easy integration of new toolsets, processes and systems, many of which will be completely unknown and unpredictable at the time of integrating current capabilities (such as Facebook, Twitter, etc) into government systems. In this light, it is more important to design our IT systems to be open to accommodate new

unexpected capabilities, rather than to provide a shopping list of semantic or social graph capabilities to select from.

3. Allow for "multiple conflicting developments" – demonstrations to try out many ideas and see what works. It for to easy to over-emphasize analysis and planning when both are of limited effectiveness.

These 3 themes are only partially addressed in this paper. A second document will deal with the much more challenging concept of social graphs and social sets ("… the heart of Facebook …"), especially from a dynamical perspective. A third document is planned for criteria and constraints of IT systems to achieve openness, flexibility, reliability, and robustness.

Definitions of Semantics

From [Dictionary.com 11Aug11], selected definitions of semantics are:

se·man·tics /sɪˈmæntɪks/ [si-man-tiks] noun (used with a singular verb)

1. *Linguistics .*
   a. *the study of meaning.*
   b. *the study of linguistic development by classifying and examining changes in meaning and form.*
2. *Also called significs. the branch of semiotics dealing with the relations between signs and what they denote.*
3. *the meaning, or an interpretation of the meaning, of a word, sign, sentence, etc.: Let's not argue about semantics.*
4. *general semantics.*

Word Origin & History - semantics

"science of meaning in language," 1893, from Fr. sémantique (1883); see semantic (also see -ics). Replaced semasiology (1847), from Ger. Semasiologie (1829), from Gk. semasia "signification, meaning."

Semantics certainly did NOT arise from text searches, rather it has long been a fundamental area of study for languages and communication. While spelling and grammar-syntax "checkers" function very nicely in applications such as word processors, email programs, it is RELATIVELY easy to build effective pattern-matching systems to do this. Semantics has been a much tougher challenge.

Rather than belabour the point, I've dug up an old "hobby-test" that I did on a "Confabulation Theory", which is mentioned in Appendix A, and shown in detail in a separate paper.

Semantic applications to the workplace – random thoughts

Semantic tools are finally becoming more familiar to many of us, but mostly in the context of more advanced "text" search capabilities such as the Google search engine. That application is easy to understand, "highly visible" to most users, and is an area where benefits can easily be generated. What Google actually does to achieve the results it gets is not generally known nor appreciated.

However, it is a mistake to limit thinking to that one, relatively simple application of semantics, because other opportunities for semantics are potentially much more profound, beneficial, and revolutionary. A very short list of "arbitrarily selected" semantic applications is:

1. **Ghostwriter © -** Assisted thematic development, information gathering/ analysis, and composition of reports "… a giant step beyond word processing …". I have not followed up on the progress of this particular concept/ product from 4 years ago, which is based on Confabulation Theory (see Appendix A). However, irrespective of the toolset, this is a general capability we should be anticipating, perhaps in gradual steps, over the next decade or so. The ability to bring new employees or outside consultants into an "organizationally specific stylistic framework" for writing could be a huge benefit (often stylistic writing and formatting is an important part of the final product, and is difficult to guess at for outsiders). It also might make a huge difference when experts are not available, but when someone "familiar" with a subject is available to take a first shot at a subject/ issue/ request. (GhostWriter is a copyright of Fairfax Corp).

2. **Organizational memory beyond records -** The gradual displacement of paper by electronic records (databases, documents, emails, spreadsheets etc etc) has not really gone much beyond the WAY that we used paper records (other than for some databases, which have required enormous resources to maintain and often to translate "records"). But the electronic availability of "native format" electronic files, combined with fast, advanced (including semantic) searches, provides us with the ability to treat available information much more like we use our own memories – a trend already well illustrated by web search engines for a couple of decades. The highly linked, and well-indexed web content, with quick, effective general sources like Wikipedia, allows us to come up to speed so fast that its almost as if we had already learned / remembered the material. Strangely, "Knowledge Search" on NRCan has only been available for about a year, and it's reach is extremely limited. Addressing information security in an effective, easy, scalable manner (from individual employees through to public-to-the-world) is essential for this to reach its potential.

3. **Checks on Organizational Filtering for Senior Managers & Employees -** This theme may fit more with the upcoming paper on "social graphs and social sets". Many problems arise in organizations because of dysfunctional message passing up and down the organization. This could become easily apparent with the right toolsets (perhaps more the social graph dynamics than semantics).

4. **Dynamic (real time) tools for [pertinence, salience, priority] of information flows** (emails, Facebook-like postings, Twitters, blogs, documents, wikis, webpages, etc). We're drowning in ever-greater information flows, so it's essential that automated tools be applied to help us keep focus and stay productive, rather than make the situation far worse. To some extent, more specialized social media environments will help, but we need MUCH more than that. In the same way that we have SPAM filters, we have to have tools to "hide" (information hiding as a valuable feature!) items that we can't afford to wade through – as executive assistants do for managers. Or, another way of looking at this is an "Un-Spam" filter that extracts a CONTROLLED flow (eg 30 items max per day) of material from diverse information flows each day.

5. **Organisational Diagnosis -** of [individual, group, organizational, network] x [dysfunction, normal-function, how-function, hyper-function] x [employee, line management, direction, regulation, client] - This capability is likely to be a complement to traditional and social [graphs, sets] analysis, but in combination with the latter may be particularly useful for rare event detection & tracking. This diagnosis would be static in the simplest analysis, where it would be a complement to traditional tools. But the only way to ??????????? It's interesting to note the portion of psychologists involved in analyzing social media applications for the government. Their background could provide a rich and powerful boost to the computer science way of looking at IT systems.

6. **Emotional analysis** (is the Mona Lisa smiling?) – is so important to verbal and visual communications, and inter-personal relations in general. A good example is the popularity of face-to-face meetings and video-conferencing as opposed to conference calls. Another example is the challenge of interpreting emotions across cultural and linguistic groups, or even across organizational cultures. The alter example of emotional content is the problem that many experienced with the onset of email and now with texting, that a lack of visual and verbal emotional signals can lead to a misinterpretation of intent or motivation, and a flaring up of communications. NRC has been looking at emotional analysis tools, which may become a useful tool for security applications. If these tools work out over time, this could also be a big boost for employee morale surveys throughout an organization.

7. **Employee surveys -** My suspicion is that feedback and surveys often miss the "real problems", as individuals are reluctant to stick out, or to being recognized as the source of complaints and criticisms. Several of the applications discussed in this sub-section (in particular the emotional analysis point above) might be useful inputs into a system that might (at a low level requiring very careful interpretation!) compliment traditional surveys, or provide feedback in the absence of surveys.

8. **Employee performance appraisals -** As pointed out in an article distributed to SPINE team members by Ken Hart [Marlow 14Jul11], social media provide a completely new, and in some ways potentially superior, form of hard data for appraisals. Whether privacy or other concerns block its use for this application is another story, but in the same way that security agencies find social media to be information rich for their purposes, performance appraisals could potentially be enriched by automated processing / analysis associated with social media tools. In a sense this has long been accomplished via Bayesian statistics for Microsoft's

internet gaming! The ranking of players and teams is critical to providing a challenging and fun onlie experience. Maybe we need to match people at work like this too?

9. **Where's Waldo**? - This concept includes rare event detection (eg trying to track down Osama bin Laden), or the automated extraction of "unusual and innocuous anomalies" that may be of importance to an [individual, group, organization], and which will generally be buried in a torrent of "noise" and distracting information.

10. **Finding what you weren't looking for -** When walking sown library isles looking for a book, it occasionally happens that we see a title on a completely different subject that turns out to be even more important than what we are looking for. It is impossible to wade through the torrent of potentially important information on the internet which may be of importance to us, but given our past emails, documents, reports etc, a system probably could recommend "strange topics" to look at. Amazon's suggestion of books to read is at one end of the spectrum. Your friend or teacher's suggestion of new people or themes is probably at the other end of the spectrum of intuition.

11. **Birds of a feather: Automated coalescence and dissolution of groups around automatically generated new themes -** It seems "natural" (?) that systems will evolve to the point that new themes can be automatically "created", and initial diverse sets of people be invited to discuss or act on different aspects of that new concept. Concepts might be technical, social, market, political or other themes, or cross-disciplinary views of an event or subject. This occurs now through the actions of organizers or advocates,, and it may be that such systems would best work through individuals and organizations who tend to do this sort of thing anyways, and who can screen and combine suggestions in creative ways.

12. **New pathways - Automated development or suggestion of new processes -** It is one thing to apply semantics to themes and issues, and quite another to apply it to processes. But we all tend to work in silos, and having a "huge world" view of the function or dysfunction of processes, and a comparison against other organizations would be a great help – especially to employees buried in a dysfunctional system and looking for supporting arguments to effect change! Consultants and auditors do this type of "forensic work", which is a check on how management has defined the organization. But I am not aware of the toolsets, and whether powerful semantic tools are being combine with hyperlink and business flow (MIS) capabilities

13. **Automated abductive logic suggestions for [individuals, groups, organizations, issues or themes]** – This is like Chrystia Chudczak's "web-page recommendations" suggested as a SPINE feature across views, but pushes an underlying, under-exploited concept.

14. **Bursting past the "frames problem" - Semantics and logic versus connectionist approaches, and hybridizing the two -** I'll not elaborate on this for now, as it "leaks into" much larger themes… and will be described in the paper related to social graphs and social sets.

15. **Contract/ agreement contradiction analysis -** Most of the time we pay senior managers and lawyers big bucks to carefully peruse major contracts and agreements. Still, everybody is extremely busy, and not everything gets that kind of attention. Going beyond grammar checking to the point where:
    - inconsistencies and could be automatically flagged for checking;
    - automated recommendations are provided with respect to an agreement's format or approach;
    - different terms and conditions could be suggested;
    - special clients for whom special conditions and constraints apply would be checked.

    Note that these capabilities could be generically applied to many other processes.

16. **Live, real-time conversations in foreign languages -** While text-based semantic applications have their own challenges, real-time translation between human languages is harder yet, partially because of semantic challenges as well as stylistic issues, and the tight performance constraints of human conversation. At one time, on my spare time I was selling fully-vocabulary voice-recognition systems to doctors and lawyers, allowing them to input directly into Microsoft Word and other applications. The obvious thing to do would be to chain the voice-recognition to a real-time translation (partially assisted with the editing features already in the voice dictations interfaces which could be adapted to a dual purpose!), and then use computer-generated speech (actually, decent computer voices are one of the biggest technical challenges). Of course, "everybody knows" that the technology isn't there to do this, but I suspect that what everybody knows is way behind what the technology can do, even if only for niche applications. Verbal communication between two individuals who don't have a common language doesn't work very well either… probably way worse than what machine assisted verbal (or text) communication has been able to do for some time..

17. **Descriptions of the styles and functioning of groups, managers, employees -** There is no end to the number of simple classification systems that people and specialists dream up. But these are arbitrary and generic. What might be the effect of un-biased, automated classification systems, that don't reveal confidential, private information, and which do not give definitive answers, but which might point us in interesting directions, and could potentially avoid some of the problems with traditional, rigid classifications? This might be even more interesting when applied to the DYNAMICS of changing organizations – in terms of personnel changes, re-organisations, major shifts in plans and the type of work or themes,

18. **Presidential audio tapes -** While only Presidents and such could afford full audio tracking of everything they say and hear, with real-time voice recognition with semantic learning and clean-up, you too could have the same service for all that you see and hear at work. Just think – you could be at the center of your own Watergate! More seriously, this would potentially be a vast, high-quality source for semantic processing and helping us to do our jobs, and remember exact details of conversations even 5 minutes ago. This could also be vastly superior to minutes of meetings. In a sense, I had tried to partially do this in the past by tying a high-quality electronic voice recorder to voice dictation software ("pre-semantic" version). Those systems have since been available for many years now, but I don't know how accurate and noise-robust they are.

The list above is far from complete, and is not intended to be a prioritized selection of semantic applications. There will be orders of magnitude bigger and better ides out there on the web and in development right now. It would be useful to cull the results of a survey looking for such concepts – but not to have a complete list or to identify THE ideas of the future: more to have a more diverse feel for what might happen.

Broad considerations

Note that an important feature of several "Semantic applications to the workplace " above is the AUTOMATED capabilities that should arise, partly based on semantic capabilities. We are used to the automated processing of accounting information, or process control systems (historically PID analogue or digital controllers), but we have not progressed as far with free-format information (text, images etc) in their native format. We should expect more and more of this at ever-more-advanced levels of sophistication.

Humans can't see and compare, in detail, any more than a handful of other individuals. In contrast, semantic processing COULD potentially cover every single individual, and all "tracked" interactions between them (the "all-seeing eye") AND it could process, abstract, and act on that information. Furthermore, it could potentially spit out high-level abstract results and recommendations WITHOUT violating privacy! (see "Crisp Privacy" below for a different type of challenge). Even a poor job with that kind of access and detail could go far beyond what any normal human could ever hope to do, even we could all do a vastly superior analysis for ANY individual that we know well enough (typically limited to peers and subordinates of our own work teams, plus a few other individuals).

Some people (I'm thinking of myself here) have emotional IQ's less than zero, and have NEVER developed a talent for analyzing and second-guessing people. Organisational attempts to develop people typically offer "conceptual frameworks" that provide simple rules, personality types, and behavioural recommendations for employees, project leaders, and managers. However, "Semantic applications to the workplace " as described in the last sub-section offer the chance to provide group or organizational-wide analysis, with advice or development for people as a potential side-effect. Perhaps more importantly, these systems might offer a basis for rejecting the "politically-correct theories of the day", for far more substantive, group-specific analysis and results. Again, that is very hard to do conventionally, so expectations will likely have to be low to begin with. **Perhaps "automated advisors and personal trainers" will arise from this.**

Is "Crisp Privacy" dead? With personal email and folders inaccessible to automated indexing and semantic processing (leaving social sets and graphs aside for now), the privacy of individuals is a given as long as access is strictly limited. With the shared directories and applications, that is less the case. However, as soon as some individuals open their email and social media type sources, then this "containment of privacy" becomes problematic, even if thematic domains are not open. Anyways, something to think about – more so when organizational change, behaviour, dysfunction and exceptional performance tools are implemented.

Leading into mid-to-long-term plans and expectations

Some of the "semantic applications" will likely be in use within a year or two in their simpler forms, others may require 5 to 15 years before they become commercially available, and a few applications may take a much longer time to mature. Others still may exist only in a hybrid man-machine form for a very long time. ***But even if a tool is not immediately available, it is still important to be aware of the potentially arising capabilities and the requirements of their underlying systems*****,** which may help orient current development and implementation projects, especially with respect to issues described in the separate paper "SPINE – Systems design issues for social media".

Endsection
****

Part II - Semantics: Technologies, and Toolsets

The relatively simple Part I provided a random sample of potential applications that semantics might help to enable. This Part II is far from complete or even from being coherent and finished as one would normally expect of a document. That is because it is only a "dipping of the toes" into the thinking in this area. I've opted for point-form listings, rather than generic, broad statements, as having the best chance of stimulating substantive follow-on thinking, including by myself.

Keep in mind that this entire document is largely from my own perspective and background, rather than being a thorough review and prioritization / comparison of semantic technologies and their significance. But there are some advantages to being a somewhat personal viewpoint, in the sense that independent viewpoints have a good chance of contrasting the consensus and conventional, raising questions that might otherwise be missed, and discouraging readers from taking an overly simplistic view of a subject, and preventing them from believing that all issues have been settled and that the "truth" has been found (a common flaw in scientists thinking – this is a key "hobby theme" of mine).

The concept lists are not sufficiently explained to clarify the concepts for readers who are not familiar with them, but I felt that it was important to provide "place-holders" for future follow-up, and for important conceptual context, rather than limit content to a few brief descriptions of selected themes. As an example of a "modest" incomplete conceptual introduction, the separate paper "Confabulation Theory – 'Plausible next sentence' survey" is 31 pages long, and cannot be considered to be anything but scratching the surface of Confabulation Theory and its relation to semantics.

Issues related to systems conceptualization and design (such as Technology versus user needs drive, unpredictable evolution and rapid roll-over of platforms, organizational and toolset

diversity etc) are addressed in the separate paper "SPINE – Systems design issues for social media".

II.1 Semantic-based office tools, processes and relationships

Classical pre-semantic tools

One can already do important analysis of text information with classical tools as described below:

- Text search – Fully-indexed text retrieval systems that work with "native-format sources" **("native-format sources" ~ [documents, emails, spreadsheets, databases, presentations, etc]**),
- Relational databases – While these have usually had extremely limited capabilities (from a user perspective with user data), have not used native-format sources, nor have they been oriented towards or compatible with unstructured information. However, there have been interesting text capabilities implemented within relational databases.
- Spreadsheets – are really the only "programming environment" for all users at NRCan, both with respect to the standard Excel functions, and with respect to the Visual Basic programming environment. There is a fair amount of power in Excel, and combined with fully-indexed text retrieval, you can do a lot.
- Spell check & Grammar-checkers - Spell checkers and Grammar checkers are important, in that they illustrate an evolution of thinking away from classical theories, such as Noam Chomsky's "rule based" grammar towards more a simpler, more general and powerful "pattern matching" pragmatic approach. That's not to say that a naturally-evolved system can't reflect underlying rules, nor is it to say that logic is an adequate framework for any such system.

Recent innovations with semantic tools

- **Google and semantic search engines -** Google had search capabilities/ effectiveness far beyond anything else on the market long before semantics became a buzzword. I have no idea of the actual algorithms used by Google, but as the dominant (only?) application of semantics that most people seem to be aware of is for better-quality searches, Google provides a worthwhile benchmark. I've forgotten the name of a more recent search engine that supposedly beats Google in some respects (not Bing!!).

- **Data-mining tools -** Earlier versions of data mining tools emphasized "word vector" matches, some of which were extremely long lists of key words, extracted automatically from general text using various statistical or other techniques. My own exposure has been in relation to the application of neural network and evolutionary computation tools to data-mining, with special attention being paid to objective functions and ?I forget the criteria's basis????....

Examples - (?IBM-Cognos ?)

It seems to me that data-mining toolsets are under-represented in many of the reviews and articles I've seen in relation to social media toolsets, and I'm not sure why?

- **Latent semantic analysis** - This particular approach is a focus of Hamid Boland and others. *[Howell – I'll leave the fill-in of details for this and other approaches for a later version of the paper]*

Missing pieces of the puzzle

Note that there are counterparts to many points below in "Part I - Semantic applications to the workplace – random thoughts".

- **Saliency, priority & "Un-spam" filters** - We are all drowning in rivers of information every day, even if we can "tune" it to all be relevant to our job responsibilities. How can we "filter this" down to a manageable amount according to the saliency, priority, and quality of the information? My guess is that "random sampling" and "re-iteration over time of key points" will be a critical part of any successful system to do this. (see also "Random Math" in •"Appendix F – Dead Math, Random Math, Live Math, and Social Math").

- **Rare event detection/ tracking** - (see "Where's Waldo? " in Part I) Some objectives require picking out non-obvious clues from an ocean of data flowing by. For example, the search for Osama bin Laden is a task for which key information will be coded or will require inside knowledge to appreciate its significance. Meanwhile a great deal of the flow of information is false or erroneous, perhaps by deliberate action of the "enemy". Another example (this time from case based reasoning) was NRC's project to help the diagnosis of maintenance problems of aircraft. Tracking down causes of a rare event by analysis of related failures and analysis of coincidental status and activities worldwide is a key capability.

- **Important novelties** - I am not aware of concepts and examples that have an ability to pick out new and evolving [themes, situations, etc] that will be important to an individual or organization via automated environmental scans.

- **Cheating theory and game theory** – Normal symbolic logic is for "dead systems in stationary environments", and therefore is a poor fit for living systems or human systems, which by nature tend to be non-stationary (unpredictable) and competitive. This is partially explained in "Appendix F – Dead Math, Random Math, Live Math, and Social Math". However, we are familiar with the application of expert systems and their descendants to games like chess. The use of semantics for understanding or composing human-language questions and responses is a first step, but what about the role of semantic tools for the formulation and evolution of strategy and tactics in more open environments? Is there more to this?

A major example of this type of system might be the "strategy games", made popular by Sim City, but also used seriously as training and practice tools by the military. How do/can advanced semantics play a role with these systems?

- **General knowledge bases -** As per my comment on "data-mining" tools above, My suspicion is that even now massive "general knowledge or domain-specific knowledge bases" are being built that will have a huge impact on the effectiveness of semantic and related tools application to real domains.

- **Power tools for building specialized knowledge bases** - My guess is that fast, effective tools for building specialized knowledge bases will become critically important. Presumably this will be able to automatically cull through an organisation's information in native format. The ability to quickly correct/ edit/ augment this will be especially important. Tools for identifying inconsistencies and gaps in knowledge will be especially important as well.

- **Structure semantics -** Information in spreadsheets and databases has a structure beyond the semantics of the text within them. Some of this may also deal with hierarchies and other specialized organizations and structures.

The following concepts will not be elaborated here, as they fit best with a separate paper on "Social graphs and social sets".

- **Approximate dynamic programming and systems self-optimization -** This whole area seems to be much better suited to connectionist approaches than symbolics, but as with the example of Confabulation Theory, perhaps hybrid approaches will have particular advantages, and there is no reason to expect that symbolic-only approaches won't have niches (large niches perhaps, because they tend to fit the "way we like to think that we think" – fuzzy systems (versus neural networks) are like that).

- **Machine consciousness -** While this subject seems completely wild, there have been very interesting, if somewhat limited, results in this area, and the need for this capability becomes MUCH clearer with some of the advanced capabilities mentioned in this paper.

- **Persistent Individual Cognition** – How often could we use the knowledge and expertise we once had in school, university, or a previous job, but the details and even some of the main principles are fuzzy or gone? Can we provide "metnal memory/cognition crutches" that help overcome this, perhaps leveraging somewhat off our previous knowledge to fill in gaps quickly? Or can tools built to assist "naïve users" be built, and serve both needs?

- **Organisational Cognition -** With the more advanced tools, there is a whole range of cognition to consider:
    - *collectivity of individual experts* – this is what we have now, including general knowledge, expertise, polices, guidelines, business processes etc.

- *Organisational cognition* (not individual) – this may become the "real Organisational cognition", as a separate entity to thee "collectivity of individual experts"
- *Hybrid individual/organisational cognition* (REQUIRES Org cognition to start, except in the trivial sense, or for very "local" events/issues/themes

This puts a whole different meaning behind "organisational knowledge and culture"!

II.2 Fashions & trends in semantics

The following listing is a sampling of my own exposure to semantic concepts over the years – it is certainly not an exhaustive list, nor am I at all up to date. Unfortunately, for this version of the paper, I am only putting in placeholder comments. Not that I remember it all anyways…

Grammar of Noam Chomsky, Non-procedural programming, Expert systems

It seems that in semantics, as with grammar, much earlier work was done using "logical frameworks and rules". But in the case of grammar checkers, pattern-matching approaches seem to work better and faster, and are far easier to build and maintain. That also seems to be a trend for semantics, including translation assistance systems. But far beyond that is to integrate connectionist and evolutionary systems approaches, which will be covered in a paper "Social graphs and social sets", although some of these techniques are listed in "Appendix E - Connectionist & Evolutionary Concepts".

Confabulation Theory (see Appendix A) is a great example of getting away from traditional approaches.

Communications, media and society (Marshall McLuhan)

While not directly related to semantics, I do feel that McLuhan's ideas on the media, communications and society are very pertinent. My own feeling is that "contentless communications" not only exist, but are becoming fairly important (this isn't a McLuhan idea, but it's McLuhanesque…). The activity levels within social graphs and sets are in themselves important, and in a crude way are exemplified by the following of the directions and volumes of encrypted information flows in wartime.

Much more on this later…

Semantic processing - An easy metaphor and example

While the appendices delve into a few more esoteric (less relevant) examples, perhaps a simple vision can be built through abductive reasoning. (*Here I use "simile, analogy, metaphor" as a working definition of "abductive" logic as a concept that belongs together with inductive and deductive logic. At least one "official" definition presents quite a different concept, and that is*

VERY different from the Wikipedia definition at http://en.wikipedia.org/wiki/Abductive_reasoning, so be warned.)

Think of a semantic processor as an extension of the historical evolution of:

• Text processors • Word processors • Voice processors with general vocabulary voice recognition (I had a Sony device several years ago)	• Flat file database • Relational database • Spreadsheet	•	• Email • Calendars • Blogs • Meetups • Wikis • Social graphs & sets
• External information – references media etc	• Employee knowledge, expertise, education	•	•
Relational database implementations of Business Models Search engines – on the web, pathetic lack of progress on WANs & organizational systems News alerts – mass media or business information flow | | V Automated semantic processing -> alerts, analysis & recommendations -> actions Uses, but also free and independent of, relational databases, although that could be one of the basis or components of the “semantic systems”.			

Enterprise-level financial accounting, inventory, personnel, sales systems, management reporting and many, many other systems based mainly on relational database technology incorporate “business models and toolsets” to do a huge amount, variety and complexity of processing of all types of information. These systems automatically summarize, process, and abstract information to produce results of the type desired.

Relational dbs – more than spreadsheets – often a mistake! Locality of information, individual capability of building tools that will never arise from the formal frameworks.

In the same manner, by going one more step we should expect “Semantic processing” to also incorporate native format text, spreadsheet, database, email, social media, presentations and video, and high-level reports or updates, and to automatically process very general, very high volume data, information, and knowledge into pertinent, salient, abstracted information and recommendations, plus taking actions and augmenting a “knowledge basis automatically!

Perhaps the best examples are in the areas of security, espionage, and (IBM-Cognos maybe?) industrial intelligence? But these are likely to be "invisible" to us.

Endsection
****

Part III – Examples of Semantic capabilities, and other concepts

I provided brief descriptions of several semantically-related examples in the Appendices:

- Appendix A - Example Game - "Plausible next sentence"
- Appendix B - Example theme - Gary Marcus' "Kludge" and "The birth of the mind"

- Appendix C - Example project - Teaching Aibo
- Appendix D - Example project – Revamping the educational systems in Africa

The intent is to make the concepts introduced in Part I and II more "concrete" for the reader.

The remaining two Appendices cover topics which are not obviously related to semantics, but I do feel that they are important issues to be aware of, in order to avoid a far too restrictive (symbolic) view of semantics:

- "Appendix E - Connectionist & Evolutionary Concepts" relates very strongly to the area of "Social graphs and sets", which will be a separate paper. But connectionist and evolutionary concepts are also of great relevance to the area of semantics. It may be useful to keep this "link" in mind. This Appendix will hopefully prevent you from adopting a framework for semantics and the social media that could easily far to narrow (?dangerously narrow?) in light of capabilities that are out there.
- "Appendix F – Dead Math, Random Math, Live Math, and Social Math" - is a shot across the bow just to make sure that you don't actually believe that symbolic thinking is the be-all-end-all basis for thought!!! There are hints of this with Confabulation Theory as introduced in Appendix A, but as Confabulation Theory is in a sense a symbolic model, one might mistakenly assume that only symbolics are involved in thinking.

As examples of key issues that may be overcome by connectionist and approaches like Confabulation Theory:

- Binding problem - "… The famous *binding problem* (von der Malsburg 1981) does not apply to confabulation theory because each of the attribute description symbols of an object is typically linked to many of the others pairwise by *knowledge links* (see Sect 1.2.2). In effect, a mental world object *is* any reasonably large subset of its pairwise-linked attribute description symbols. …" [Hecht-Nielson 2007, p8]
- Frames problem - "… I'm missing the quote here …" [Lendaris and Santiago 2005]

References

[05Sep2011 Howell – incomplete, missing some information for ~5 or so references. Missing papers for data mining, of which I reviewed several some years ago. The listings below are dominated by Connectionist themes, which are not the central theme of the paper.]

[1] Tom Cobb 2010 Private conversations. University of Montreal

[2] Dictionary.com 11Aug11 "Semantics" http://dictionary.reference.com/browse/semantics

[3] Facebook developers reference site ("… the heart of Facebook …") ?hyperlink

[4] David .B. Fogel, (editor) Evolutionary computation: The fossil record: Selected readings on the history of evolutionary algorithms. Piscatoway, NJ: IEEE Press, 1998 Excellent primer on evolutionary computation

[5] David B. Fogel, Blondie 24: Playing at the edge of AI. San Diego, USA: Academic Press, 2002. p301 for comments on learning and evolution Awesome lesson on concepts much closer to profound machine learning than classical AI/ expert systems type of approach

[6] Walter J Freeman 14Jun2009 "Cortical Mechanisms of Memory Formation and Readout" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009 p882-889 Extremely interesting (and important) discussion of perception, black noise and memory. A separate paper in a post-conference workshop covered synchrony issue.

[7] Donald Hebb of McGill ~1940's or 50's? - Hebbian learning (couldn't find refrence quickly…)

[8] Robert Hecht-Nielson 2007 "Confabulation Theory: The mechanism of thought" ISBN 978-3-540-49603-8 Springer-Verlag Berlin Heidelberg 245pp, accompanied by computer DVD for learning

[9] Guang-Bin Huang, Xiaojian Ding, Hongming Zhou 09Feb2010 "Optimization Method Based Extreme Learning Machine For Classification" Accepted by Neurocomputing, 22pp In this paper, Guangbin and team show how ELMs subsume support vector machines. If correct, this is huge!

[10] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks," GMD - German National Research Institute for Computer Science, GMD Report 148, 2001, http://www.faculty.iubremen.de/hjaeger/pubs/EchoStatesTechRep.pdf.

[11] Edgar Koerner 14Jun2009 "Brain-like Intelligence Inside - Towards Autonomously Interacting Systems" IEEE Proceedings of IJCNN 2009, International Joint Conference on Neural Networks. Atlanta, paper#0700, Plenary talk

[12] Marshall McLuhan 1962 "The Gutenberg Galaxy: The Making of Typographic Man" University of Toronto Press. Actually, many of MacLuhan's concepts that are most relevant to Social Media are in other papers and books. I need to go back to those sources for the next version of this paper

[13] Gary Marcus 2008 "Kluge: The haphazard construction of the human mind" Houghton Mifflin Co., Boston, New York21pp

[14] Gary. Marcus, The birth of the mind: how a tiny number of genes creates the complexities of human thought, New York: Basic Books, 2004. This book (along with Pinker's "Blank slate") is an essential read, and its concepts and Marcus' current work are a basis for the current paper.

[15] Iain Marlow 14Jul11 "Boosting the performance of performance reviews" Interview of Mr. Debow at the company Rypple, about Loops software, Globe & Mail (as pointed out by Ken Hunt 15Jul11) http://www.theglobeandmail.com/report-on-business/managing/on-the-job/boosting-the-performance-of-performance-reviews/article2097832/page2/

[16] John S. Mattick, "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms" BioEssays, vol 25 pp930-939, Oct. 2003 widespread in conceptual impact - opens the vista on the potential role of non-genetic DNA

[17] Marshall McLuhan ?year? "The Guttenberg Galaxy" ..

[18] M.J. Meaney, M. Szyf, "Maternal care as a model for experience-dependent chromatin plasticity?". Trends in Neurosciences, Vol.28 No.9 September 2005.

[19] Leonid Perlovsky - Neuro Linguistic modeling (reference…)

[20] Leonid Perlovsky "Computational Intelligence: Natural Information Processing" tutorial 70$US http://www.ieee.org/education_careers/education/ieee_course_catalog/view_course_catalog/course_details/index.html?COURSE_ID=735 He refers to cognition and language in this.

[21] Steven Pinker 2002 "The Blank Slate: The modern denial of human nature" Penguin books, 2003 edition 509pp I did not complete the entire book, but a good chunk
[22] Danil Prokhorov 31Jul2005 "Echo State Networks: Appeal and Challenges" Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, pp 1463-1466, 31 July - 4 August 2005.
[23] Roberto Santiago, George G. Lendaris, "Reinforcement Learning and the Frame Problem", Proceedings of IJCNN 2005, International Joint Conference on Neural Networks. Montreal, paper#1659, pp 2971-2976, 31 July - 4 August 2005.
[24] John Taylor 2006 "The Mind: A users manual" John Wiley & Sons, Chechester, West Sussex 286pp My favourite model of consciousness - much like engineering control or game theory
[25] Jean-Philippe Thivierge 05Dec2008 "Synchronization in the Brain: Deciphering the Development, Damage, and Recovery of Neural Interactions" This seems to be similar to a presentation Thivierge gave at NRC in Ottawa in ~2008-2009 period that I attended (he's now at NRCC, Ottawa)
[26] I.C.G Weaver, N. Cervoni, F.A. Champagne, A.C. D'Alessio, S. Sharma, J.R. Seckl, S. Dymov, M. Szyf, M.J. Meaney, "Epigenetic programming by maternal behavior". Nature Neuroscience vol 7, no8, Aug. 2004, pp847-854 great paper, in ~Feb2009 a Globe and Mail article mentioned he was getting private funding for seeing if childre from disturbed / abusive families might be helped in this way, What about psychological trauma of soldiers?
[27] Paul .J. Werbos, "Neurocontrol and supervised learning: an overview and evaluation", and "Approximate dynamic programming for real-time control and neural modelling". in D.A. White, D.A. Sofge (editors), Handbook of intelligent control. New York: Van Nostrand Reinhold, 1992. This is a classic that introduces the framework for later work on ADP, Much more mordern reviews were done by Danil Prokorov, Don Wusch, and others
[28] Bernard Widrow 14Jun2009 "Cognitive Memory: Human and Machine" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009, p3365-3372 Bernie's plenary re-iterated "memory molecules" as REQUIRED for long-term memory, Most hated it, but a mid-career great neuroscientists who does the wet work, said that Widrow is correct in stating that long-term memory CANNOT be sustained via synaptic modification, which is the current "scientific religion"
[29] Lofti A. Zadeh 23Jun2003 "Computing with words and perceptions (cwp)-a paradigm shift in computing and decision analysis" IEEE Proceedings of IJCNN 2003, International Joint Conference on Neural Network, Portland Oregon (this may have been an impromptu talk?)

endsection

Appendix A - Example Game "Plausible next sentence"

A.1 Background for the Game/ Exercise

My own feeling is that the common perception of semantics is limited to its application to improving the capabilities of search engines. This is a fantastically myopic view of semantics! I felt that a simple exercise based on "Confabulation Theory" would give an easy, very powerful feel for semantics and its potential and significance.

Confabulation is a biologically-inspired theory for cognition that was developed by Robert Hecht-Nielson of San Diego. It is described in:

- a brief overview in the next sub-section,
- somewhat greater detail in the paper "Confabulation Theory – 'Plausible next sentence' survey" based of work on my own account in Calgary during the period 2007-08, and
- Robert Hecht-Nielson 2007 "Confabulation Theory: The mechanism of thought" ISBN 978-3-540-49603-8 Springer-Verlag Berlin Heidelberg 245pp, accompanied by computer DVD for learning

As per [Howell 2008]:

> "… As a gross summary, Confabulation Theory assumes that information is held within "attribute classes" in roughly 4,000 thalamocortical modules (carrying information about "mental object attributes") and roughly 40,000 cortical knowledge bases (establishing "meaningful co-occurrences" between thalamocortical modules). All vertebrates (and even invertebrates such as bees and octopi) are postulated to possess functionally analogous structures, albeit in smaller quantities. Confabulation is a "winner-take-all" process for coming to a conclusion (intermediate or final) , and is the only information-processing operation used in cognition. Confabulation DIFFERS from Bayes theorem in statistics, and these simple differences make confabulation a superior form of reasoning for the real world, where information is often incomplete, erroneous, or event misleading (predator – prey). It is even proposed that many supposed successes of Bayesian statistics are the result of extreme simplifications which mean that it is actually Confabulation that is being applied, without the statisticians and scientists even being aware of this important distinction! …"

Confabulation Theory provides a great example for the theme of this paper ("Semantics beyond search"), which is to force a broader context of thinking with respect to the potential impact of advancing semantic capabilities on the toolsets and processes at work, and especially for the implementation of social media systems in the workplace.

While the "Next Plausible Sentence" exercise doesn't directly illustrate semantic processing in the workplace, it WAS a critically important eye-opener for me, and gives a very strong "hands on" sense that while technology has a very, very long ways to go, it is a very, very long ways further along than commonly thought.

But Confabulation Theory goes much further than just semantics. It may be right, it may be wrong, or it may be irrelevant: it is still an amazingly powerful, insightful, and fun look at one concept for our "Mechanism for Thought". As quoted in Part I of this paper, Confabulation Theory: "… presents the first concrete and detailed (and thus falsifiable) scientific theory of how thinking works. …"

And to repeat here a semi-quote of the author of the theory back in 2002:

"… Our advice to you is to start your research immediately,
To run as fast as you possibly can, and to never look back.

In a few short months, you will hear the starter's pistol firing behind you,
Unleashing the greatest intellectual land-rush of all time. …"

[Hecht-Nielson ?2002? at the World Congress on Computational Intelligence, Hawaii]

Try the exercise on your colleagues, and have fun!

A.2 Confabulation Theory – an overview by quotes of the conceptual basis

This exercise was taken from results presented in the excellent book:

Robert Hecht-Nielson 2007 *"Confabulation Theory: The mechanism of thought"* ISBN 978-3-540-49603-8 Springer-Verlag Berlin Heidelberg 245pp, accompanied by computer DVD for learning

I had first seen the concept of Confabulation Theory a few years earlier, during a presentation / dramatic announcement by Robert Hecht-Nielson at the Word Congress on Computational Intelligence in Hawaii, 2002. While the book and DVD is a solid and necessary starting point for understanding Confabulation Theory, several quotes from the book are provided here as a flavour of the theory (in blue font), and to show that it is anything but trivial.

The descriptions of Confabulation Theory blow are selected and quoted directly from Hecht-Nielson's book as cited above. While it is common practice to re-interpret material from another author, I feel that this is inappropriate, and that the reader is best served by quotes than (mis)-interpretations. (If you interpret this as laziness on my part, well that's true too…).

[p 5] "… **Cerebral Cortex and Thalamus: the seat of cognition**
…
This book presents the first concrete and detailed (and thus falsifiable) scientific theory of how thinking works. This confabulation theory proposes the specific neuroanatomical structures, and their functions, that are involved in human cognition.

The two main human neuroanatomical structures postulated by confabulation theory to be involved in the implementation of thought are thalamocortical modules (figure 1.1) and knowledge bases (Fig. 1.2). These structures, which constitute the "information-processing hardware" used to carry out thought, exist within the cerebral cortex and thalamus. The human brain possesses roughly 4,000 thalamocortical modules and roughly 40,000 knowledge bases. All vertebrates (and even invertebrates such as bees and octopi) are postulated to possess functionally analogous structures, albeit in smaller quantities. …"

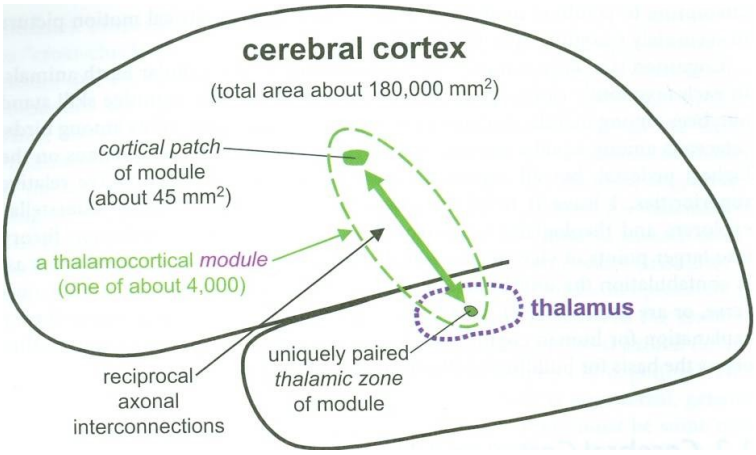

Fig. 1.1 A *thalamocortical module* (one of roughly 4,000 in the human brain). Each thalamocortical module is comprised of a small *patch* of cerebral cortex and a uniquely paired *zone* of thalamus. The cortical patch of each module is reciprocally axonally connected with the thalamic zone of the module. The cortical patches of different modules are largely disjoint (partial overlaps do likely occur). Similarly for their thalamic zones. The union of the cortical patches of all thalamocortical modules comprise the entire area of the cerebral cortex. However, the union of the thalamic zones of all modules do not comprise all of the thalamus.

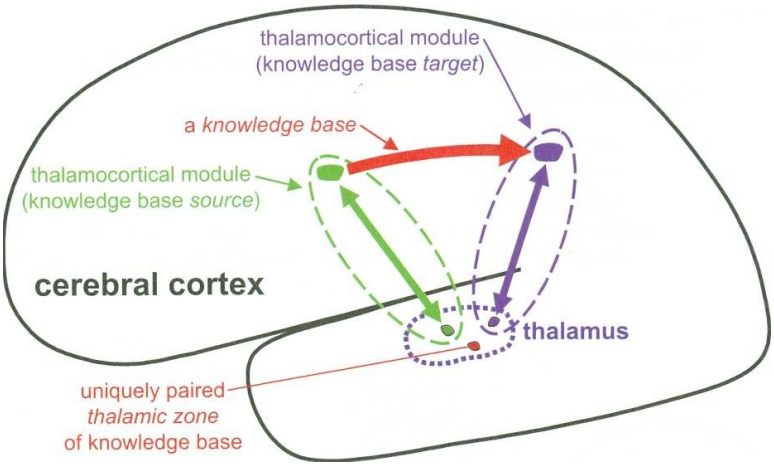

Fig 1.2 A cognitive *knowledge base* (one of roughly 40,000 in the human brain). Roughly 40,000 ordered pairs of thalamocortical modules (source and target modules) are selected (by genetically specified developmental processes carried out in childhood) to each have their cortical patches unidirectionally linked by a knowledge base. Each knowledge base is comprised of a large number (often millions) of individual *knowledge links*. Much like a thalamocortical module, each knowledge base is postulated to be paired with a unique, dedicated zone of thalamus which is postulated to be involved in that knowledge base's *enablement*. The combination of the thalamic zones of the modules and knowledge bases make up the vast majority of the thalamus.

A.3 A brief description of the Confabulation system for "Plausible Next Sentence"

Hecht-Nielson's "Confabulation Team" built an exercise around using two sentences to generate a "third plausible sentence. The Confabulation system itself was fed ?120 million sentences and 70 million sentence pairs? from quality newspapers, magazines etc. These were "serious sentences and sentence pairs" - not jokes, plays on words or other sources that may intentionally obscure meaning. Spelling, grammar, syntax were of good quality in the information used as the basis for their system.

Several tests of the "next plausible sentence" generation by confabulation theory are provided in the book [Hecht-Nielson 2006?], and it is from those examples that the current exercise was drawn. Note that these test sentence examples were taken randomly from quality newspapers, magazines, and similar sources. They were NOT part of the training examples for the Confabulation system. So the latter had to "invent" new responses to new questions, so to speak. Those responses also required correct spelling, grammar, syntax, which is an outcome (side effect?) of confabulation theory.

The result of the overall survey, was that only one out of approximately 20 "Execisees" (those who judged the Respondents) figured out who the "affected respondent" was, that is, the respondent "… who had difficulty in understanding and composing sentences / text …". Exercisees were given two hints in succession, the last of which revealed that the "affected respondent" was a machine. In other words a "machine" or system that was never given any rules of spelling, grammar, syntax, was able to autonomously build not only a third plausible sentence that "looked normal" (albeit it usually ranked in the lowest ¼ to 1/3 of respondents by the quality of its responses).

However, there real test and surprise was that NO "Exercisee" immediately spotted (within 30 seconds of reading the instructions and Question #1's responses) that one of the respondents was totally out of whack, which is what I would have expected with the technology that I was aware of during that era (2002-2006)!

endsection

Appendix B - Example theme - Gary Marcus' "Kludge" and "The birth of the mind"

[Marcus 2008]

Gary Marcus is a linguistics professor at New York University, who did his PhD under well known linguist Steven Pinker. Both are closely associated with Noam Chomsky. His book "Kludge" challenges the mainstream idea in linguistics that language is a near-perfect communications capability, arguing instead that it is a Scotch tape and bubblegum bag of tricks that makes the best of what is available with existing cognitive processes and systems. What Marcus does not state, but which is obvious, is that if language capabilities are immature, there will likely be huge increases in capabilities, perhaps though large step changes over time. Evolutionary step-changes in human capabilities, genetically or epigenetically anchored or not, cannot be excluded. Presumably, technologies could potentially help drive this, but I doubt that people would be so comfortable with the thought?

Appendix C - Example project - Teaching Aibo

[Koerner 14Jun2009]

Edgar Koerner is (was?) head of Honda Research Europe's advanced robotics lab, which is responsible for the high-end cognitive development of the famous Honda Asimo robot. It has long been learned that one cannot simply program robots, that the process of learning is fundamental to higher level functioning. Note that David Fogel (see the list of references) has claimed that evolution is also a NECESSARY requirement of higher level learning, and he makes a good case of that with his checkers and chess playing systems.

Appendix D - Example project – Revamping the educational systems in Africa

[Cobb 2010]

Tom Cobb is a linguistics professor at the University of Montreal who has recently become involved in a project to assist with an African country's educational systems. While not purely semantic, some of his expertise may help make an impact in that area. While I have not the time to elaborate on this, it is one small example of the potential applications of semantics to the area of education.

endsection

Appendix E - Connectionist & Evolutionary Concepts

As with Appendix A, below is a sampling of my own exposure to concepts over the years which may or may not have an application to semantic processing of various types – and again, it is certainly not an exhaustive list. The list is specifically focused on Computational Intelligence tools – and not the "whole universe" of techniques that are useful for semantics. Subjects for which I took a tutorial, attended a special session/ plenary, or have discussed /emailed first-hand (all of the proceeding with the founder himself) are indicated by an asterix (*). Obviously, I tend to reflect my direct interactions with individuals, rather than having an overall view of activity in the field.

I have provided only a single reference for each concept, as that is all that I have time for now. As explained earlier in this paper, the themes will be covered more thoroughly in the paper on "Social graphs and sets".

My own guess is that the toolsets are at an extremely immature stage, and that toolsets will arise quickly based on radically new completely outcomes from the "machine learning" field. These will likely involve different "philosophies of thinking", and not just more effective extensions of scientific thinking, or of the current connectionist and evolutionary concepts.

E.1 Computing with words [Zadeh 23Jun2003]

E.2 Hebbian learning [Donald Hebb, McGill ~1940's or 50's?]

The information theoretic dereviation of Hebbian learning by Harold Szu of the US Office of Naval Research was a hugely impressive bit of work – easy for Szu, but important in the implications for information theoretics and to give a sound explanation of why Hebbian learning from psychology is so important, and is still used today.

E.3 Neuro Linguistic modeling [Perlovsky]

E.4 Network reconnection & brain recovery from insult [Thivierge 05Dec2008]

E.5 Echo State Networks [Jaeger 2001]

E.6 Extreme Learning Machines (ELM – * Huang Guangbin, Signapore)

E.7 Bayesian statistical theory versus Confabulation (?Michael Rogers of Microsoft?, * Robert Hecht-Nielson)

E.8 Swarm Intelligence (I forget the founder's name – I've met him a couple of times) and **Collective Intelligence** – as in robotics (Mikalainen's Plenary at IJCNN 2006 in Vancouver)

E.9 Stochastic chaos, synchronization, & the brain [Freeman 14Jun2009]

E.10 Memory molecules [Widrow 14Jun2009]

E.11 Lamarckian heredity, the brain + behaviour switching/ modification [Meaney, M. Szyf, 2005] [Mattick 2003]

> *"… The objective of our group is to destroy the central dogma of all biology – genetics. …"* [Mattick 2003]

E.12 Consciousness [Taylor 2006]

Concepts NOT included in this appendix:

Quantum computing
Autonomous mental development ()
Everything else…

Note the theme from an IEEE Computational Intelligence Society (IEEE-CIS) call for papers on the theme of the Semantic Web. Again, the brick wall that traditional approaches often meet is being addressed here by Connectionist and Evolutionary techniques. However, it isn't always the fanciest and most powerful techniques that work!

==========

3.5 CFP: IEEE Computational Intelligence Magazine: Special Issue on Semantic Web Meets Computational Intelligence
Posted by: Kay Chen Tan (eletankc@nus.edu.sg)
Date submitted: August 15th, 2011

IEEE Computational Intelligence Magazine
Special Issue on Semantic Web Meets Computational Intelligence

Call for papers

The Semantic Web (SW) carries out the vision of a global web of data usable for both human and machines. This web, consisting of inter-connected instance data annotated with possibly expressive ontologies, promises a huge opportunity for web-based applications in many domains. As an open, decentralized, and complex system, the building and development of the Semantic Web bring up a number of new problems to which traditional Artificial Intelligence approaches (Description Logic, Logic Programming, Ontology Reasoning, etc.) are not effective and feasible. More and more researchers from the Semantic Web community are aware of the importance of introducing computational intelligence methodologies and approaches (e.g., including neural network, fuzzy logic or evolutionary computation) to study the complexity of

such a huge web of data.

The goal of this special issue is to discuss on current trends in the marriage of the Semantic Web and Computational Intelligence. The specific topics solicited for this special issue will cover all aspects of computational intelligence research, including, but are not limited to,

- Fuzzy logic and uncertainty reasoning for Semantic Web
- Neural network and machine learning for Semantic Web
- Evolutionary computation for Semantic Web
- Hybrid approaches of EC, NN, and fuzzy logic for Semantic Web applications

All papers submitted to this special issue will be rigorously peer-reviewed by experts in the field. To ensure timely publication of this issue, the following timeline will be used for this special issue:

Manuscript Submission Deadline: October 15, 2011
Notice of Review: November 15, 2011
Final Manuscript Due: December 15, 2011

We look forward to your submissions to this special issue!

Guest Editor:

Huajun Chen (Zhejiang University, China)

endsection

Appendix F – Dead Math, Random Math, Live Math, and Social Math

> *"… The purest form of art, beauty, and sheer fun IS mathematics. …"* [Howell ~1998?]

Dead Math

Here I use the phrase "dead system" to mean any non-living/ non-human system. Scientific thinking and methods do not give rise to a "solid truth" or result for complex dead systems in the same way they do for simple systems where the preconditions for "classical" logic/science are in place:

1. We know all of the key variables
2. We can measure the variables (including "internal" system variables) precisely
3. We can control independent variables precisely
4. We can measure all key effects precisely
5. The mathematics is tractable.

For complex, dead systems that do NOT fulfill the classical pre-conditions, we get a false, or over-confident, sense of truth:

> *"… Logic is an emergent property. …"*

Even when very good mathematical/numerical models are available for a complex system:

> *"… Complex system behaviour has typically NOT emerged from even high fidelity, complex models in science. Almost invariably, the complex behaviour of systems is observed, then back-fitted into the models. In that sense the models have been quite useful, but the problem has been in projecting the image that the models give something that they do not. …"* [Howell ~2007?]

Thus is highly misleading for scientists to claim that their "scientific models" can be trusted to provide a "true" representation of complex systems

> *"… The assemblage of "true" model components, based on sound phenomenological concepts, is no guarantee that the resulting composite system is correct. In practice, these composite systems have a disastrous track record (climate models, CFCs and ozone, macro-economics, ???). Scientists were forewarned by Kurt Gödel in the 1940's (provability theorems 1 & 2) and David Wolpert in the 2010's (1 and 2), and for the most part still haven't learned this lesson. …"* [Howell 2008-2010?]

> *"… Phenomenologically-based complex models progressively approach the status of "small-world universal function approximators". As with universal function approximators, they may actually have little to say about the phenomenology, and could give a false confidence in their apparent representation of the system's phenomenology. …"* [Howell, conjecture ~2008]

Random Math

> *"… Any system capable of [modeling, predicting, controlling] sufficiently complex systems, must have as one (but not all) of its critical components, a stochastic (random) process. …"* [Howell 2010, - sic. circular definition to some extent]

Live Math

All living things, including viruses (whether or not you prefer to include them as "living") fall into this theme. Evolution, and "behaviour" (non-mental or "robotic" changes by bacteria amoebae, plants etc) are included here, as conscious thought is NOT a requisite for adaptive responses to environmental opportunities, threats, or changes!

Social Math

> *"… The reason for your success, as [an individual, an organization, a society, or a civilization], has everything to do with Cheating Theory and Game Theory, and very little to do with [Rational, Logical, Scientific] thinking. You can hire people or buy/steal their ideas to do the latter, but in competitive environments you must be good at the former. …" [Howell ?2005-2009?]*

Science – a foundation for truth in classical areas, but over-extended

"Classical" areas of science, such as basic physics and chemistry,

Endsection

enddoc