A Cognitive Informatics Framework for Language Understanding

L. Massey
Department of Mathematics and Computer Science
Royal Military College
Kingston, Canada
massey@rmc.ca

Abstract—**For centuries, science and philosophy of human language have posited that language understanding is a logico-deductive process. In this paper, we argue that this is not necessarily the case, at least not for basic forms of understanding such as the ability to determine what a text document is about. We present a human-inspired Cognitive Informatics (CI) framework for basic language understanding that relies on the semantic relationship between words as embodied in conceptual knowledge stored in Long Term Memory (LTM). Short Term Memory (STM) limited capacity, neuromorphic spreading activations and neural decay are also contributing to meaning determination. A computational model implementing the framework shows the potential of the approach by establishing that the output of the model is as good as what humans generate.**

Keywords - Language understanding; text analysis; cognitive informatics.

I. INTRODUCTION

Human language understanding has long been characterized as a logico-deductive process. Indeed, syntactic and semantic analyses, among others, rely heavily on logical structures to infer meaning. Given the lack of success and the rigidity of this paradigm, coupled with large quantities of electronic texts and increasing computational power, computational linguists have in the last two decades sought meaning in the enormous amount of co-occurrence information found in corpora. Applications of this idea can be found, for instances, in corpus-based statistical methods like Latent Dirichlet Allocation (LDA) [1], Latent Semantic Analysis (LSA) [2, 3]) and text clustering [4, 5, 6]. Document classification [7, 8] is also related to this paradigm. It involves the application of supervised machine learning techniques to approximate a classifier function based on the recognition and generalization of patterns found in a large number of training samples.

However, it is unlikely that humans achieve language understanding by relying on the presence of millions of documents in some sort of mental corpora, or by observing and generalizing on thousands of previously classified exemplars. One might claim that reproduction of natural processes or biological plausibility is not a requirement for Artificial Intelligence (AI) or Natural Language Understanding (NLU). The example often stated is that planes do not fly by flapping their wings like birds, relying instead on the principle of aerodynamics. Hence, maybe one should endeavor to similarly uncover a fundamental principle that would make NLU more effective. On the contrary, one might argue that due to the relative lack of success in that regard, biological and human cognitive processes might be a valuable source of inspiration. It is the position taken in this paper.

We introduce a Cognitive Informatics (CI) [9] framework for basic NLU inspired by how the human mind might process information when one reads. We address the same problem as corpus-based approaches to meaning and understanding, namely to seek a restricted, basic form of meaning in the form of themes present in a text document (i.e., the topics, the concepts).

Our contribution is to show that it is possible to obtain the main themes of a natural language text without relying on large corpora, training sets or rule-based and knowledge intensive parsing and semantic analysis. Instead, we propose a framework that involves a simple and elegant cognitive process based on activation and decay of conceptual knowledge stored in long-term memory coupled with a limited capacity short-term memory and neuromorphic spreading activations. We call this process ReAD, which stands for **Re**treival (of conceptual knowledge from long-term memory), and **A**ctivation and **D**ecay (of this knowledge). We note that it is possible that more in-depth understanding might require more elaborate processing. For instance, the framework introduced here may be useful in identifying the general semantic context - i.e. what a fragment of text is about - to guide and improve the effectiveness and efficiency of other language analysis that may be required for more detailed understanding.

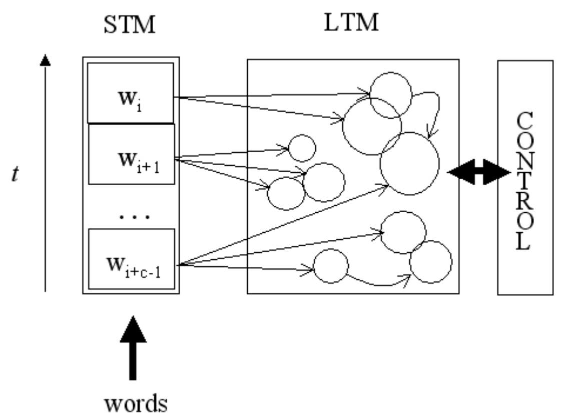

Figure 1. A Cognitive Informatics framework for basic NLU.

The rest of the paper is organized as follows: the next section describes our CI framework; section III presents an illustrative implementation of the framework, examples of processing and initial results obtained with that implementation; and section IV engages in a discussion on the framework, pointing out, along the way, to related work in CI. We conclude and identify on-going and future work in section V.

II. A CI FRAMEWORK FOR NLU

Overall, the framework (illustrated in Fig. 1) involves the sequential processing of words in Short-Term Memory (STM), access to and activation of knowledge stored in Long-Term Memory (LTM), and concurrently, decay of the activation over time. Note that the time arrow in Fig. 1 is indicative of the sequential processing of words in STM, and does not apply to decay in LTM. We call this process ReAD, which stands for Retreival of conceptual knowledge from long-term memory, and Activation and Decay of this knowledge.

The ReAD cognitive process takes place once individual words have been recognized (i.e. we do not address the issues related to word recognition). A similar process could take place for language heard rather than read, once the auditory system has decoded the sounds and recognized words.

A word w_i triggers the retrieval and activation of LTM *ideas* t_{ik} related to w_i. An idea is akin but more vague than the notion of concept [10] as a cognitive unit to represent and embody semantics of real-world objects. We use the word idea because it is more evocative of the fuzziness and imprecision one may feel until the mind comes to a conscious conclusion about the text being read.

The action of receiving a word w_i by a cognitive processing apparatus implementing the framework is thus one of:

$$w_i \text{ -> } \{(t_{i1}, a_1), (t_{i2}, a_2), \ldots, (t_{in}, a_n)\}$$

where -> denotes the access and activation of ideas tik related to wi and stored in LTM. Each tik is associated with an activation ak. This activation is independent of words, since different words can activate the same idea.

Each circle in Fig. 1 represents an idea tik with straight-line arrows indicating the activating association with a word. The size of the circle indicates the level of activation of an idea. In the brain, clearly this corresponds to neural activation potentials. LTM ideas are any sensory experience related to wi and stored in LTM. It can for instance be sounds, images, etc that evoke in a person's mind anything rational or emotive related to wi. Hence, the word related ideas in LTM correspond to the mental model for word wi an individual has constructed over time. Ideas can overlap and are related by a notion of semantic similarity (the distance between circles in the figure can serve to illustrate that point: circles separated by more distance are less semantically related). Ideas are by nature non-local and may in turn activate other ideas stored in LTM, each cascading activation [11, 12] resulting in a lower level of activation. These secondary activations are shown as curved arrows in Fig. 1 (there can also be tertiary activations, etc). There is thus a neuromorphic chain reaction of cascading activations feeding onto one another, each with a dampening factor γ proportional to the distance of the original word activation:

$$w_i \text{ -> } \{(t_{ik}, a_k)\} \text{ -> } \{(t_{kj}, \gamma\, a_j)\} \text{ -> } \{(t_{jm}, \gamma^2\, a_m)\} \text{ ->} \ldots$$

A word is stored in STM until capacity c-1 is exceeded. The limited capacity of STM is a well-known phenomenon [13] and it may play a critical role in the derivation of understanding rather than being a mere computational constraint [14]. Words in STM are constantly replaced by new ones from the text being read sequentially or from the utterance from an interlocutor. When in STM, a word maintains the focus of attention for its associated ideas, and consequently their activation level is sustained maximally. When a word wi looses the focus of attention, the activation of its associated ideas tik starts to decay [15]. However, other words may come along and reactivate an idea that would otherwise decay to a dormant state. The exact nature of decay may vary, but it is more likely exponential to account for the natural decrease of voltage potential in neurons.

Hence, different words can activate or re-activate the same ideas, resulting in increased activation levels. At any point in time, but most likely when the end of a natural language utterance or text is perceived, a control module (on the right side of Fig. 1) is triggered to focus attention momentarily on the state of LTM. This action may be performed by an attention control process such as the central executive [16], or by a consciousness mechanism that brings awareness onto ideas activation. Nevertheless, the goal of this module is to observe and interrogate the state of LTM activations to determine the most active ideas. This is deemed to be what the language fragment processed up to that point is about. Hence, highly activated ideas crystallize into the understanding of what concepts, themes or topics were present in the language stimulus. The activation level is indicative of the confidence or clarity one may have of the meaning. How many ideas are thus retained is determined by an activation threshold in the neural network responsible for the above mentioned control mechanism.

III. Implementation and Results

A. Implementation of the framework

How ideas are acquired and how exactly they are represented in the brain is beyond the scope of this paper and the topic of intense research in cognitive science. Here, one can just assume ideas exist in the mind, as clearly they do for humans. This may be sufficient for a purely theoretical framework, but we also aim at demonstrating the feasibility of implementing the framework as a computational model. Therefore, one must consider the question: how can ideas in a biological brain be emulated in a computer? Additionally, one would rather avoid the extremely resource consuming task of handcrafting all possible ideas in a large knowledge base, as often attempted in AI. Practically speaking, this is not a viable solution. This has, in fact, been a source of frustrations in AI and in NLU. Usage of statistical techniques to extract word co-occurrence information from large corpora has been a relatively profitable answer to the problem of knowledge acquisition. Nonetheless, as mentioned previously, this is implausible cognitively. Further, it is a source of its own set of problems, such as being computationally expensive and increasingly dependent on the traditional NLU techniques one initially aimed at avoiding (along with the associated knowledge acquisition issues).

Considering the above, we suggest that existing source of information on words can be used to provide conceptual knowledge to emulate LTM. This can be, for instance, dictionary definitions or encyclopedic entries. The intention is to use words in on-line dictionary definitions or in on-line encyclopedic entries such as Wikipedia as surrogates for ideas. This is obviously a very rough approximation of what would happen in a real biological brain, but nevertheless, if this rough approximation can be shown to work sufficiently well, then it would demonstrate the validity of the framework. Consequently, one would expect to obtain even better results with more cognitively plausible and rich sources of knowledge. One can think of the Web as a candidate.

The algorithm shown in Fig. 2 lists the basic steps involved in a simple instantiation of the framework with a dictionary as a source of knowledge replacing LTM ideas. The algorithm processes words from a textual document. In this case, we refer to words in the dictionary as *items* to distinguish them from words from the text. The items correspond to ideas from the framework. Stop words (such as articles, pronouns, etc) are filtered out according to a list of such words as is commonly done in text analysis. This is required since these words will not generate useful items.

```
Input: Document, dictionary

For each word wi obtained sequentially
from document:
  Query dictionary with wi
  Ti = items from dictionary
  Remove stop words from Ti
  For each item Ti = {ti1, ti2, …, tin}
      Increase item’s activation
  For each item not associated with
    current word wi
      Decay item’s existing activation

Output: The most activated items
```

Figure 2. A simple algorithm for the ReAD process.

The most activated items can be determined by a threshold of activation or specified as input depending on the requirements of a particular application or experimentation.

It is the conceptual relationships between words, as embodied in the overlapping of items, and the interplay of activation and decay that are of prime importance. This simulates the activation of ideas in LTM. Note that the algorithm in its current instantiation (Fig. 2) neither implements limited capacity STM nor spreading activation. Activation is limited to the items present in the dictionary entry. For example, a dictionary entry for word "bank" might look like this:

```
1. financial institution that
takes money, deposits and lends
money
2. sloping land beside a body of
water
3. flight maneuver in which an
aircraft tips laterally
```

The resulting list of items for word bank would be the list of all non-stop words for all senses listed above. Only those items will have their activation incremented. We started with this simpler implementation of the framework to demonstrate a proof-of-concept. It is interesting to note that there is no complex linguistic processing in this algorithm, including no word sense disambiguation: it is a simple matter of increasing and decaying items activation to determine the main concepts or topics in the document. There is also no dependence on large corpora to extract statistics or learn: the algorithm works directly on a single document.

B. Examples

We now show two actual examples obtained when processing text documents with an implementation of the algorithm described above.

Document 1:

```
Mounting trade friction between the U.S. and Japan has raised fears among many of Asia's exporting nations that the row could inflict far-reaching economic damage, businessmen and officials said. They told Reuter correspondents in Asian capitals a U.S. Move against Japan might boost protectionist sentiment in the U.S. And lead to curbs on American imports of their products. But some exporters said that while the conflict would hurt them in the long-run, in the short-term Tokyo's loss might be their gain.
```

The two most activated items produced for document 1 are:

```
capital  business
```

Document 2:

```
Sri Lanka gets USDA approval for wheat price. Food Department officials said the U.S. Department of Agriculture approved the Continental Grain Co. sale of 52,500 tonnes of soft wheat at 89 U.S. Dlrs a tonne C and F from Pacific Northwest to Colombo. They said the shipment was for April 8 to 20 delivery.
```

Topic produced for document 2: `cereal`

One can observe that a simplistic implementation of the framework proposed here results in relevant topics. The words used as topics are keywords that evoke in a potential user's mind the general nature of the document and can help, for instance, decide whether or not a document corresponds to the information need currently pursued. We note that the topics produced are not words present in the text (e.g., cereal is not in document 2), but items from the definitions. This is what is specified by the framework: the most activated ideas become the concepts representative for the language. In fact, to make that point clear, in the implementation of the algorithm we explicitly removed the words from the text from the list of items obtained for a word. If a topic keyword can also be found in the text, then it must be because it corresponds to an important concept repeatedly activated *by other words*. In the case of document 2, some highly activated items (e.g., 'letter', 'alphabet') were manually removed because they were irrelevant. Some fine-tuning of the implementation is still required to handle such useless items. Indeed, one will recall that these results were produced with a limited, rough approximation of what the framework describes i.e. the spreading activation and decay of distributed, wide-ranging, multi-modal knowledge stored in the human brain. Yet, clearly some intelligent behavior emerges from the simple algorithmic implementation demonstrated above.

C. Experimental Evaluation and Results

For a more complete evaluation, a sample of the topics obtained with our algorithm on the 3299 documents of the "ModApte" test set of the Reuter-21578 Distribution 1.0[1] were evaluated by three independent assessors. The sample of size *n* was determined to be at least 332 for a confidence level of 95% and confidence interval 5%, including a correction for finite population.

[1] http://www.daviddlewis.com/resources/testcollections/reuters21578

Documents are about economical news. Evaluators were told to entirely read one document at a time. After reading a document, they were asked to write, beside the document number in a spreadsheet, a few words (1 to 5) of their own that describe the specific topics covered by the document. They were asked to provide specific topics (for example, 'acquisition', 'money market', 'financial results' but not 'economy'). The instructions insisted on the importance of completing this step first to avoid being influenced by the next step. The assignment of their own topics was decided to firm up the opinion of the assessors before they check the algorithm output and also for comparative tests of their topics with computer-generated ones. The next step consisted in looking up the topics assigned by our program to the same document. The task was to read the words used as topics, to rate them as follows and to write the score in the spreadsheet. A score of 0 or 1 means respectively that the topics are totally wrong or wrong. The distinction between both is not that important but left room for the judgment of the assessors. A score of 4 means that the topics are exactly on target and a score of 3 means that the topics are good. A typical distinction of a score of 3 vs. a score of 4 might be that the topics are too general or abstract but still describe well what the document is about. Finally, a score of 2 is acceptable. Evaluators were told this might be when the topics are related and generally acceptable (in the sense of not being wrong), but not close enough in their opinion to warrant a score of 3. The assessors were briefed to ensure their rating was based on the document content and not only on their own topics assignment, and to go back to the document and read it again in case of doubt. The average number of documents for which our ReAD implementation attributed a topic judged acceptable to perfect according to the three human assessors was 36%.

Our next step in work currently underway will be to ask other human evaluators to score various computer-generated topics (from our ReAD implementation and other topic identification methods) and the topics that were identified by human evaluators in the experiment described above. As a first step in that direction, we have compared the human- and ReAD-generated topics with the gold standard topics provided with the Reuter dataset. We scored the topics with the F1 metric, a widely used measure of quality in text mining [17, 18]. The results are shown in Table I. One can observe that there is no significant difference between the quality of topics generated by humans and those generated by our algorithm implementing the ReAD cognitive process, even though this implementation is limited compared to the framework we described.

TABLE 1. COMPARISON OF HUMAN- AND READ-GENERATED TOPICS.

	F1
Human Topics	0.56
ReAD Topics	0.54

The general nature of a framework aims at hiding implementation issues. Accordingly, several details are not shown here, such as parameters determining the rate of activation decay and the capacity of STM. These values are implementation and application decisions, but we nevertheless point out their existence briefly here. Furthermore, one can envision several different instantiations of the framework. For example, it may be possible to implement a system based on the framework without decay of activation, or as we have done here, without cascading activations and limited capacity STM. The value of each implementation might be application specific or of a scientific nature (to measure their effect). We left these variants, implementation issues and more detailed experimental investigations for future work, rather concentrating on describing the framework itself and presenting promising initial empirical evaluations.

IV. DISCUSSION AND RELATED WORK

In the CI framework for NLU we presented, it is the timely interactions of and relationships between ideas along with the passage of time (which results in 'forgetting' i.e. decay) that cause understanding. Meaning naturally emerge due to convergence to a representative small set of ideas that stay sufficiently and determinately active in LTM. This is related to the Object-Attribute-Relation (OAR) model [19] of CI that describes how LTM is based on relations rather than the traditional memory as a container. It is also related to the work of Just and Carpenter [20] who have developed a computational theory of STM in which capacity, time and activation are found to play key roles in language understanding. They also share with our framework on the importance of activation not being just for information maintenance but a fundamental computational element. However, instead of solely engaging activation and decay, Just and Carpenter's model is implemented with production rules that propagate activation and perform syntactic and semantic analysis, clearly casting their work within the traditional logico-deductive and knowledge intensive NLU.

It is important to observe that the framework described here works without in-depth linguistic processing and functions outside the realms of conventional production rules common in AI and computational modeling of cognitive processes. In our case, meaning, disambiguation, and theme identification emerge naturally from the mere interactions of memory activation and decay over time rather than from a handcrafted set of logical rules representing linguistic and world knowledge on which deductive inferences are applied.

In accordance with the Layered Reference Model of the Brain (LRMB) [21], the framework presented here deals mostly with the intermediate layers of memory and perception. This is with the exception of the control module, which is a conscious interpretive cognitive process of the mind that interrogates memory to come to a conclusion as to what a natural language utterance or text document is about. The memory systems themselves are subconscious and autonomic [22]. Ideas activation depends on an individual's social and personal history context, as stored in the brain. As well, there are no explicit and deterministic procedural instructions other than the pre-defined process of words triggering propagating activation in the nervous system. Conscious intervention occurs only when the attention-control system interrogates LTM to determine meaning.

The framework we propose does away with cognitive processes as formal inferences and formal knowledge systems to produce meaning. This is a powerful idea when one considers the logico-deductive historic roots of NLU science and philosophy. In particular, one could say that in the case of the ReAD implementation, text basic meaning (topics, themes) is computationally derived from simple lexical information. In the wider view of the framework we propose in this paper, the same basic meaning arises from the activation and decay of ideas in the distributed, multi-modal mental lexicon. However, logico-deductive syntactic and semantic analysis may still be required for more in-depth processing. Hence, one can see the framework hereby described as a context provider to guide further inferential processing. This could offer an opportunity to improve NLU effectiveness and efficiency.

We can also situate our framework within existing theories of perception and the classic debate about the role of the stimulus versus experience. Our framework fits neither while at the same time deriving ideas from both. Indeed, on one hand, meaning arises bottom up (structuralism) [23] from simple analysis of raw sensory data (words) to ever increasing complexity of analysis. (topics, concepts).

On the other hand, from a constructivist top-down processing perspective [24], knowledge and meaning is generated from previous experiences and prior knowledge stored in LTM by making unconscious inferences about what one reads. These inferences are not of the logico-deductive kind, rather being simple neural associations that relate ideas that have aspects in common.

However, the framework proposed here is emergent and reificative in the sense that the final output is more than what can be observed in the input stimuli. Thus, one might say our framework is more compatible with Gestalt theory [25], given that 'the whole is more from the sum of its parts'. Another obervation supporting Gestalt theory is the nativist aspect of the ReAD process that presents understanding as the autonomous interactions between ideas rather than a learnt skill. This is not a rejection of learning, but certainly a reduction of its importance, rather focussing on natural emergence of structure and dependence on mental and stimuli contexts.

Finally, one cannot ignore the information processing view of cognition, and particularly the connectionist processing model [26] of Rumelhart and McClelland. This model emphasizes that information is distributed in the brain as networks of connections. In this model, activation spreads from neuron to neuron when a certain threshold in summed input stimuli is reached.

From a practical computational linguistic and corpus statistical perspective, the ReAD process replaces large collection of text by accessing a source of information on words such as a dictionary. This offers the advantages of more efficient computing and processing single text in isolation.

The framework algorithmic implementation exploits the power of term dependence arising from a document lexical cohesion. Contrary to existing text analysis methods [4-8], documents are not represented with vector-space and bag-of-words modeling. Consequently, without vectors, there is no issue with vector dimensionality. We thus embark on a representation for text that is, in fact, representation-less, in the spirit of embodied AI [27]. One may object that text must still somehow be represented. Indeed, text is represented as text itself, i.e. as a series of sequential words, but not stored otherwise than very partially and momentarily in STM. There is no need to scan through a corpus to collect distributional data on words. There is no loss of information related to the order of words within a document. In the framework proposed here, a document text is preserved in its original form as a stream of words and meaning represented as evocative keywords.

The framework and the implementation we presented are not attempting word sense disambiguation. Instead of choosing one single sense for a specific word, the goal is to look for a small set of recurring ideas corresponding to the various senses across the whole language utterance or document. These repeating patterns of ideas are then deemed to be strong thematic indicators, and can be collapsed into a representative concept by the perception module. Hence, we are not interested by the meaning of single words but rather by the global interactions of activation and decay of the ideas related to multiple meanings, leading to global meaning.

Meaning in this context is macroscopic and is about identifying the main themes directly from the content of the document by itself rather than by comparing its similarity with other documents as in clustering or by words distribution patterns as in classification. The themes or topics are represented by keywords that are not extracted from text but are a product of the activation and decay of ideas in LTM. The keywords correspond to the most activated ideas. A string representation of the underlying concept is not a necessary component of the framework, but it is a practical one.

Indeed, topics are the answer to what the document is about, presented as a list of evocative keywords. It is well understood that topics can shift along a document and that a document topic may not be unique, depending on application, context of use and perceptions. Indeed, a human user's interpretation is complementary to the computational process, even an integral part of it. Topics exist in this symbiotic relation rather than in vacuum. The topics (or again, if one prefers, the themes or concepts) present in a natural language utterance or text are abstractions of the keywords, yet the keyword itself is used to label the topic. For instance, a concept may be represented by a string label such as "animal" if the document is discussing pet care. Another concept might be represented by the string "pet".

The ideas themselves are the fuzzy and imprecise set of mental experiences related to a word. Once enough evidence has been accumulated, the ideas, through increased neural activation caused by repeated access to the same brain areas triggered by successive words, collapses into precise conceptual knowledge. That is the way the concepts prominent in the language being listened to or read are produced.

In the end, for practical communication with the real-word, the goal is to generate string labels representing the underlying concepts in such a way as to evoke in the user's mind the nature of the document and answer the question of what this document is about. Just as documents are representation-less, meaning is relying on mere strings that label concepts and on their interpretability in the user's mind. There is no complex formal representational required.

V. CONCLUSION AND FUTURE WORK

This paper introduced a Cognitive Informatics framework for basic language understanding that neither relies on a logico-deductive process nor on word co-occurrence corpus statistics. Instead, the framework proposes to exploit the sequential processing of words within a limited capacity Short-Term Memory as well as the retrieval and activation of ideas stored in Long-Term Memory. Concurrently, decay of the activation takes place over time and a control module interrogates Long-Term Memory to select the most active ideas as the main concepts present in the natural language utterance or text. We call this cognitive process ReAD for Retrieval (of LTM knowledge), Activation and Decay. The application of the process results in a basic understanding of language, namely the identification of the main concepts, themes or topics.

Interestingly, the framework posits that determination of meaning may be partly due to a simple autonomic memory system in which activation of relationships between ideas associated with words are the fundamental computational element. In the implementation we presented, meaning was derived by computing with activation of lexical information used in lieu of mental ideas. This simplified implementation of the framework was carried out to successfully demonstrate what can be achieved despite not using the full capacity and richness of a biological brain as well as the full potential of the framework. Particularly, we have shown experimentally that the quality of topics obtained with the simple computational model implementing the ReAD cognitive process is similar to the topics humans identify.

From a practical point of view, the framework offers a lighter alternative and a complementary approach to existing language analysis systems and semantic technologies for ontology learning [28, 29], automated semantic tagging and cross-referencing [30, 31], knowledge acquisition from text and from the web [32-34], and information management (for instance, text classification/clustering, or more generally information retrieval).

We have implemented the framework in a proof of concept software system with which we are presently conducting further empirical evaluations using the web as a human brain emulator and comparing with other topics identification methods [35, 36]. We are also working on refining and formalizing the framework. One exciting research endeavor will be to implement the framework neuromorphically or at least in a more plausible neural manner with a richer and more complete set of experiential ideas. As well, the investigation of the actual neural reality of the framework using brain imaging [37] would be a valuable scientific endeavor.

REFERENCES

[1] D. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.

[2] T. K. Landauer, and S. T. Dumais, "Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge." Psychological Review, 1997, 104 (2), 211-240.

[3] D.S. McNamara, “Computational methods to extract meaning from text and advance theories of human cognition.” Topics in Cognitive Science, 3(1), pp. 3-17 (2011).

[4] A. K. Jain , M. N. Murty , P. J. Flynn, “Data clustering: a review.” ACM Comput. Surv. 31, 3 (1999).

[5] R. Feldman , J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data (Cambridge University Press, NY (2006).

[6] L. Massey, “On the quality of ART1 text clustering.” Neural Networks, 16, 5-6 (2003).

[7] F. Sebastiani, “Machine learning in automated text categorization.” ACM Comput. Surv. 34, 1 (2002).

[8] X. Qi, B. D. Davison, “Web page classification: Features and algorithms.” ACM Comput. Surv. 41, 2 (2009).

[9] Y. Wang, “The theoretical framework of cognitive informatics”, Int. J. Cognitive Informat. Natural Intel., vol. 1, pp. 1-27, (2007).

[10] M. R. Quillian, “Semantic memory”. In M. Minsky (Ed.), Semantic information processing. Cambridge, MA: MIT Press. (1968).

[11] A.M. Collins and E.F. Loftus, “A Spreading-Activation Theory of Semantic Processing”, Psychological Review, Vol. 82, pp. 407–428, 1975

[12] P. Van Den Broek, “A 'landscape' model of reading comprehension: Inferential processes and the construction of a stable memory representation”. Canadian Psychology Psychologie canadienne, Vol 36(1),pp. 53-54, Feb 1995.

[13] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information.” Psychol. Rev. 63 (1956).

[14] W. Schuler, S. AbdelRahman, T. Miller, L. Schwartz, “Broad-coverage parsing using human-like memory constraints.” *Comput. Linguist.* 36, 1 (2010).

[15] J. S. Nairne. “The Loss of Positional Certainty in Long-Term Memory”. Psychological Science, vol 3, pp. 199-202, May 1992.

[16] A.D. Baddeley. “Working memory”. Oxford: Clarendon Press, 1986.

[17] M.F. Hussin and M. Kamel, Document clustering using hierarchical SOMART neural network. In International Joint Conference on Neural Networks(IJCNN), 2003.

[18] L. Massey. “Evaluating and Comparing Text Clustering Results”. In Proceedings of 2005 IASTED International Conference on Computational Intelligence, 2005.

[19] Y. Wang, “The OAR model for knowledge representation.” In Proceedings of the 19th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE’06) (pp. 1696-1699), Ottawa, Canada (2006).

[20] M. A. Just, P. A. Carpenter, “A capacity theory of comprehension: Individual differences in working memory.” Psychol. Rev. 99 (1992).

[21] Y. Wang, Y. Wang, S. Patel, & D. Patel, “A layered reference model of the brain.” IEEE Transactions on Systems, Man, and Cybernetics (Part C), 36(2), pp.124-133 (2006).

[22] X. Jin and J. Liu, From Individual Based Modeling to Autonomy Oriented Computation, in Matthias Nickles, Michael Rovatsos, and Gerhard Weiss (editors), Agents and Computational Autonomy: Potential, Risks, and Solutions, pages 151–169, Lecture Notes in Computer Science, vol. 2969, Springer, Berlin, 2004.

[23] Gibson, E. (1969). Principles of Perceptual Learning and Development. New York: Appleton.

[24] Gregory, R, L. (1970). The intelligent eye. Jarold & Sons Ltd: UK.

[25] Köhler, Wolfgang (1947): Gestalt Psychology. New York, Liveright.

[26] Rumelhart, D., & McClelland, J. (Eds.). (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge, MA: MIT Press.

[27] R. A. Brooks, “Intelligence without representation”. Artificial Intelligence (47) pp. 139-159, 1991.

[28] A. Maedche and S. Staab, “Ontology Learning for the Semantic Web.” IEEE Intelligent Systems 16, 2 (March 2001), 72-79.

[29] A. Zouaq, M. Gagnon and B. Ozell, “Semantic Analysis using Dependency-based Grammars and Upper-Level Ontologies”, International Journal of Computational Linguistics and Applications, 1(1-2): 85-101, 2010.

[30] D. Milne and I. H. Witten. “Learning to link with wikipedia”. In Proceeding of the 17th ACM conference on Information and knowledge management (CIKM '08). ACM, New York, NY, USA, 509-518. 2008.

[31] H. L. Kim, S. Scerri, J. G. Breslin, S. Decker, and H. G. Kim. “The state of the art in tag ontologies: a semantic model for tagging and folksonomies.” In Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications (DCMI '08). Dublin Core Metadata Initiative 128-137, 2008.

[32] L. Massey, “Contrast Learning for Conceptual Proximity Matching”, International Conference On Machine Learning And Cybernetics (ICMLC 2007), 19-22 Aug 2007, Hong Kong, 2007.

[33] W. Wong, "Learning Lightweight Ontologies from Text across Different Domains using the Web as Background Knowledge". Doctor of Philosophy thesis, University of Western Australia, 2009.

[34] T. M. Mitchell, J. Betteridge, A. Carlson, E. Hruschka, and R. Wang, "Populating the Semantic Web by Macro-Reading Internet Text", Invited paper, Proceedings of the 8th International Semantic Web Conference (ISWC 2009), October 2009.

[35] L. Massey and W. Wong, "A Cognitive-Based Approach to Identify Topics in Text Using the Web as a Knowledge Source", In: Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances; IGI Global (Accepted for publication, July 2010).

[36] L. Massey , “Autonomous and Adaptive Identification of Topics in Unstructured Text”, In: Proc. of the 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Kaiserslautern, Germany, 2011.

[37] M. A.. Just, V. L. Cherkassky, S. Aryal & T. M. Mitchell, “A neurosemantic theory of concrete noun representation based on the underlying brain codes”. PLos ONE, 5(1), 2010.