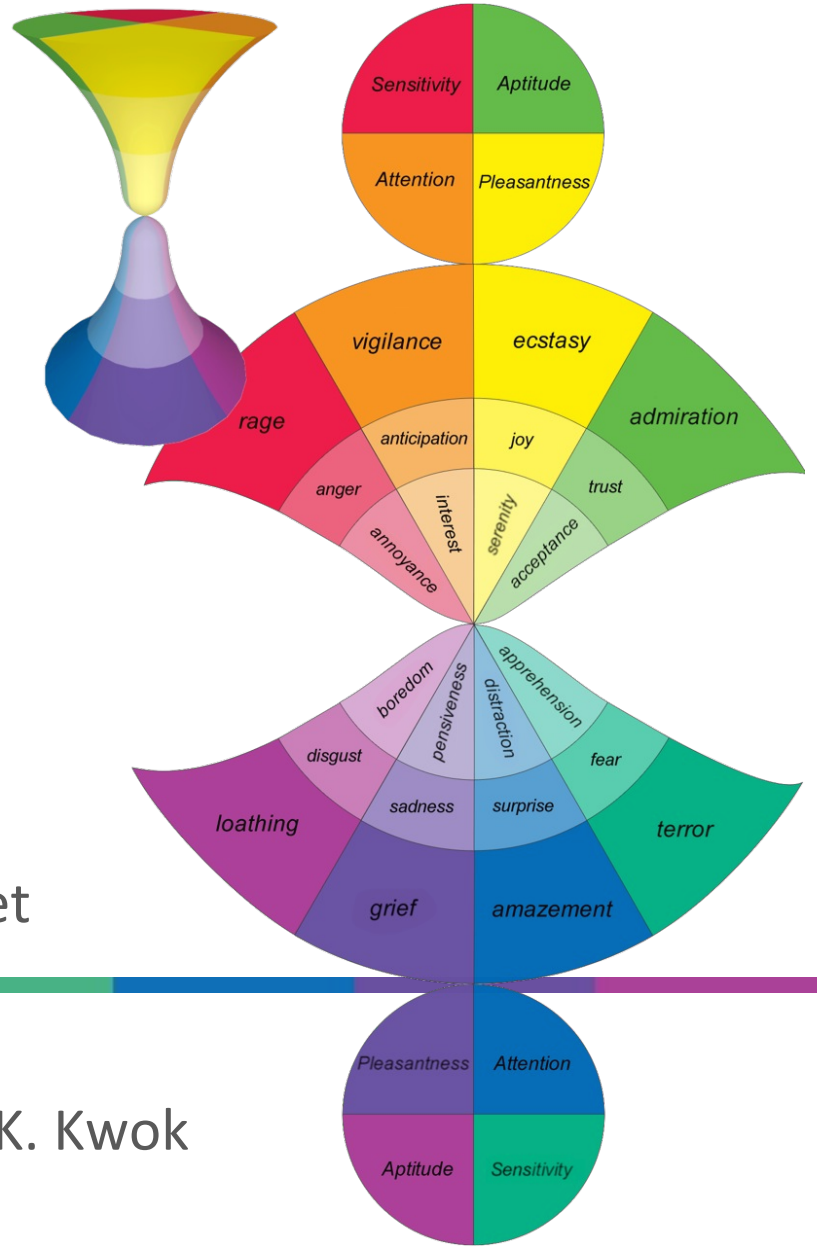
A Graph-Based Approach to Commonsense Concept Extraction and Semantic Similarity Detection



Skype: senticnet
Web: http://sentic.net
Email: cambria@media.mit.edu
Twitter: http://twitter.com/senticnet
Facebook: http://facebook.com/senticnet

13th May, 2013

D. Rajagopal, E. Cambria, D. Olsher, and K. Kwok
Cognitive Science Programme, NUS

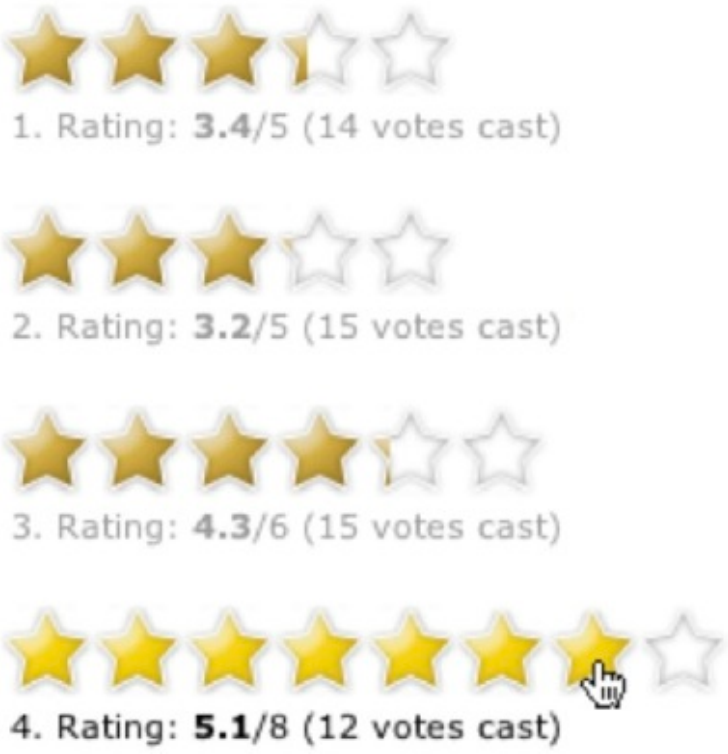
Collected Intelligence

Information today is extremely portable and processable. However, this collected intelligence, as mainly based on natural language, is far from being able to be addressed as collective intelligence.

[1] T. Gruber. Collective Knowledge Systems: Where the Social Web meets the Semantic Web. Journal of Web Semantics 6(1), pp. 4-13 (2007)

Not So Structured

According to different evaluation schemes, a very positive and a very negative review can have the same star rating. Star systems, moreover, do not assess products and services on a feature-based basis.



[2] M. Hu and B. Liu. Mining and summarizing customer reviews. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Seattle (2004)

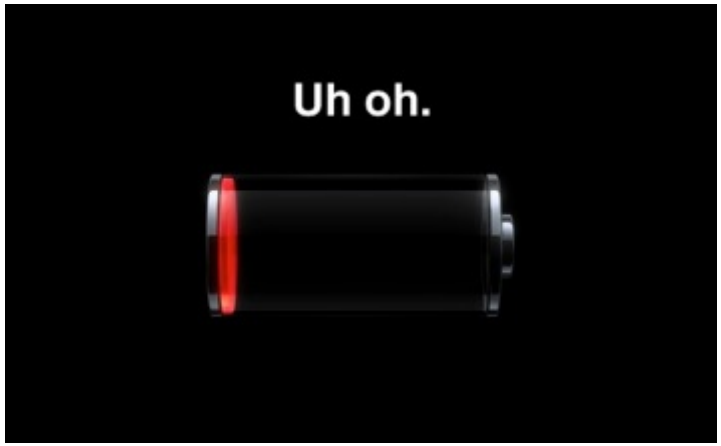
Sentiment Analysis

Sentiment analysis research, a recent yet popular and growing field, evolved from heuristics to discourse structure, from coarse- to fine-grained analysis, from keyword- to concept-level mining.

[3] E. Cambria, B. Schuller, Y.Q. Xia, C. Havasi. New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, doi:10.1109/MIS.2013.30 (2013)

Feature-Based Analysis

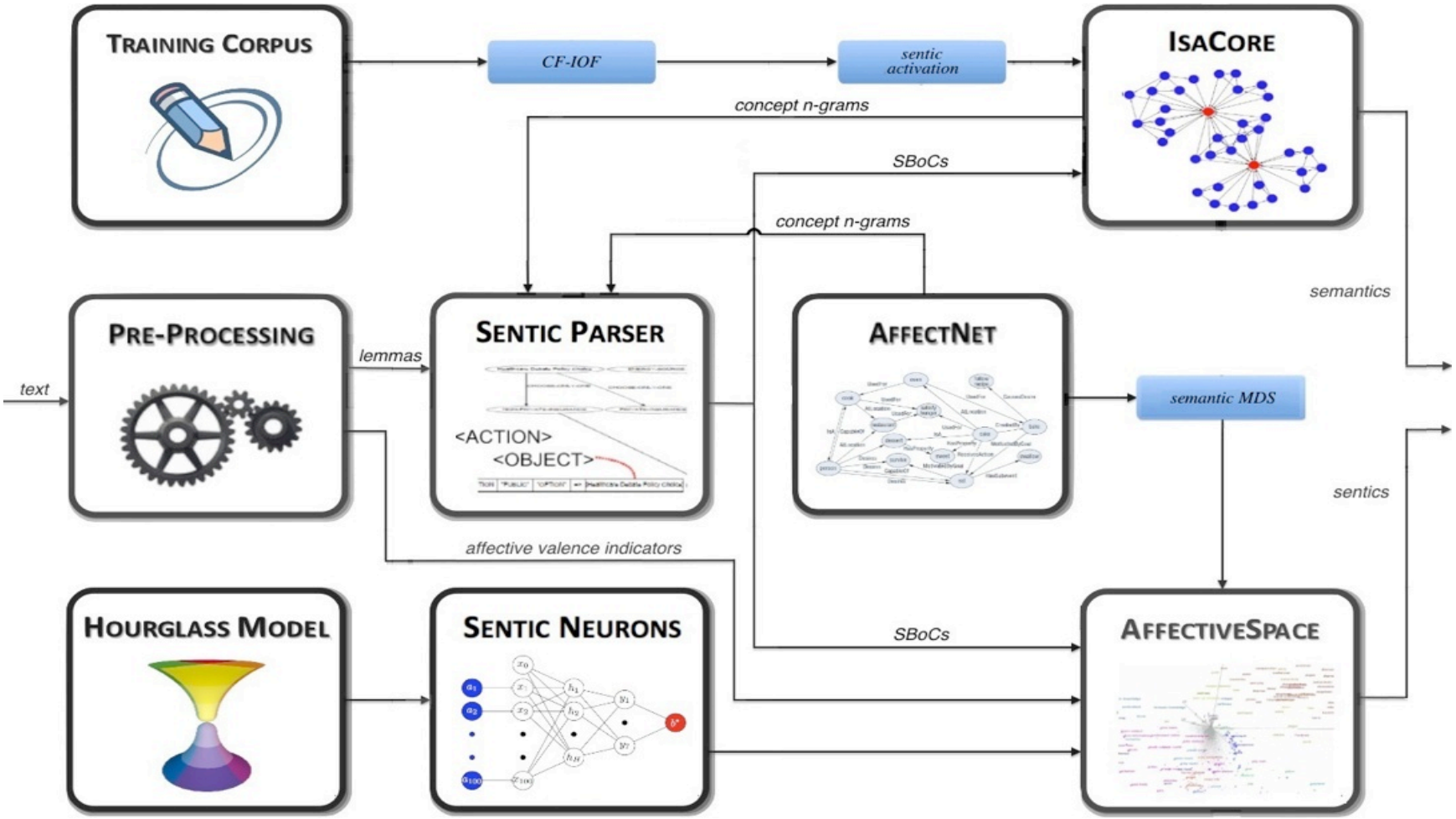
I love the new iPhone5 screen! the battery life sucks though…



document/paragraph-level approach: neutral polarity

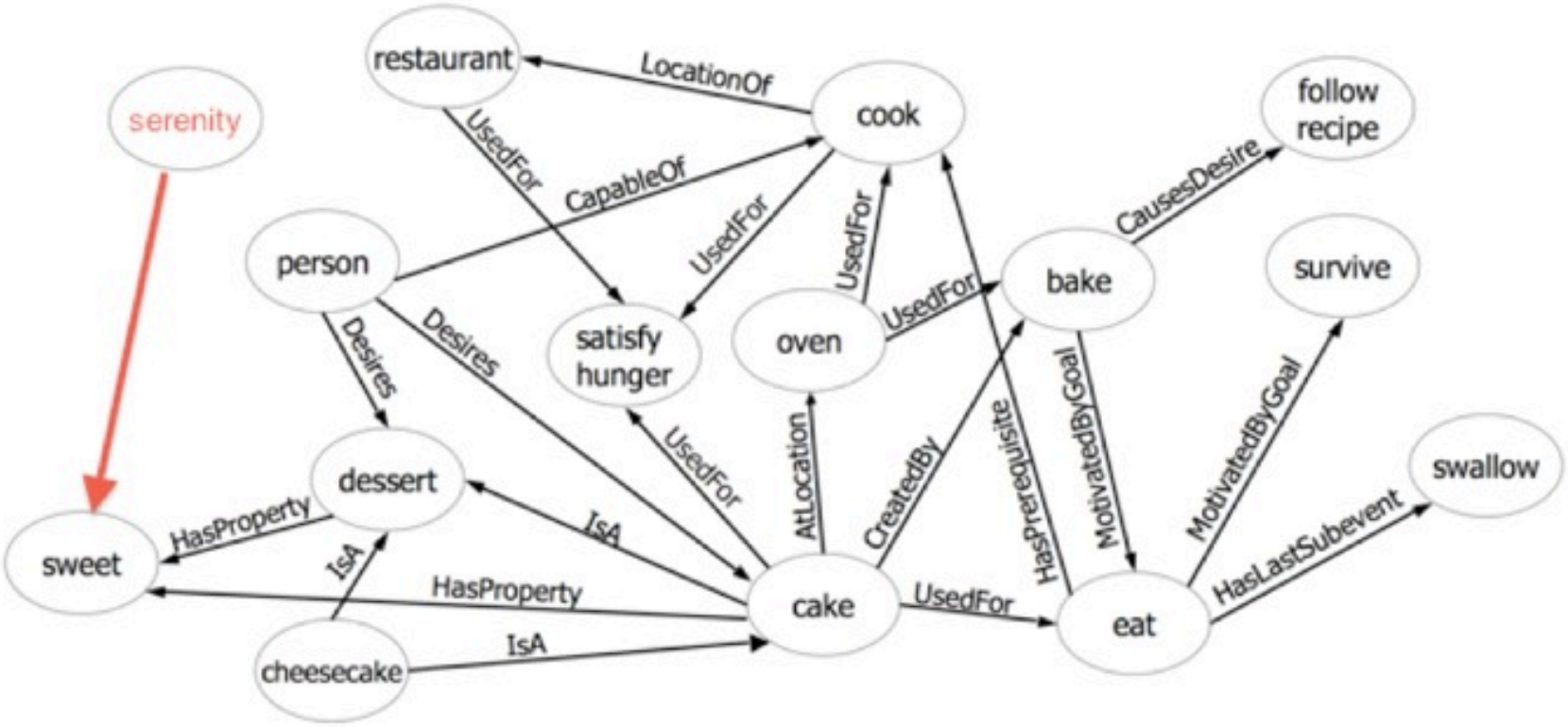
clause/concept-level approach: screen+, battery-

Sentic Computing



[4] E. Cambria and A. Hussain. Sentic Computing: Techniques, Tools, and Applications. Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2 (2012)

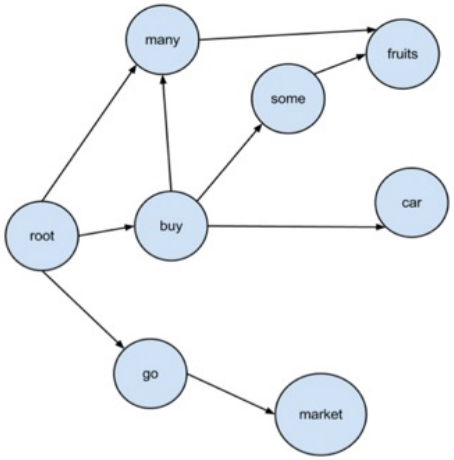
AffectNet Graph

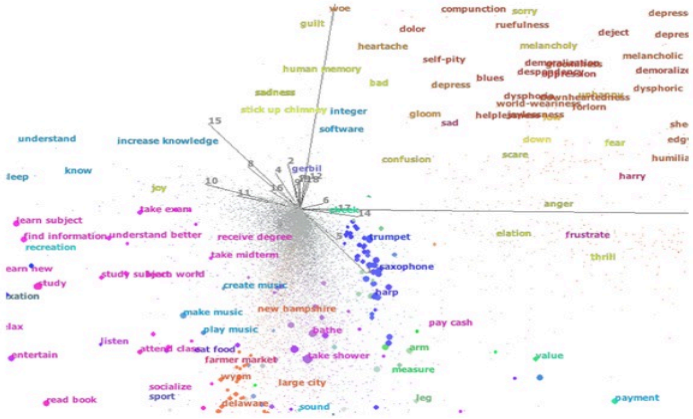


[4] E. Cambria and A. Hussain. Sentic Computing: Techniques, Tools, and Applications. Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2 (2012)

Sentic Parser

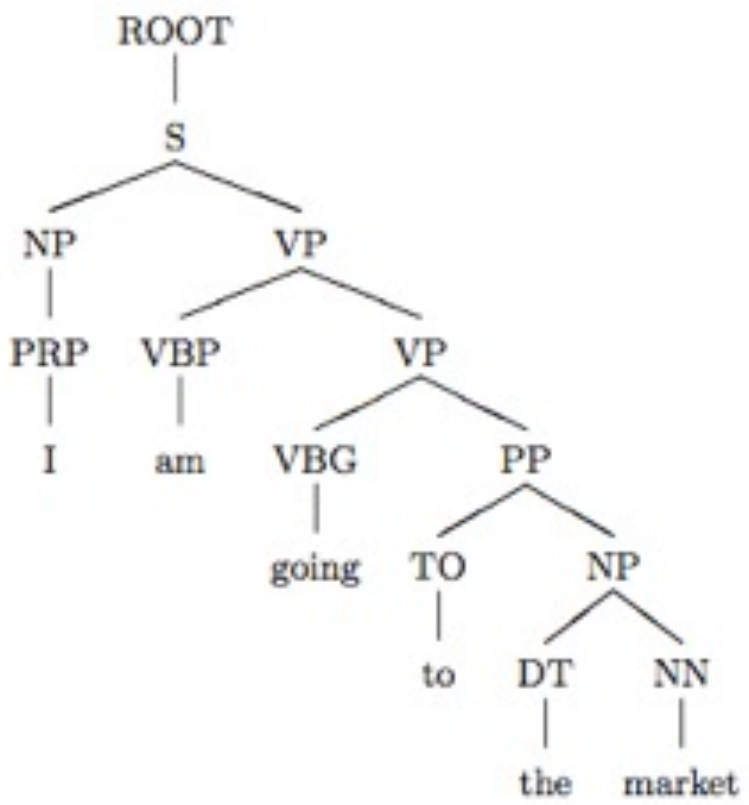
The main aim of the Sentic Parser is to deconstruct text into concepts. To this end, a graph-based concept extraction algorithm and a MDS-based similarity detection technique are hereby proposed.

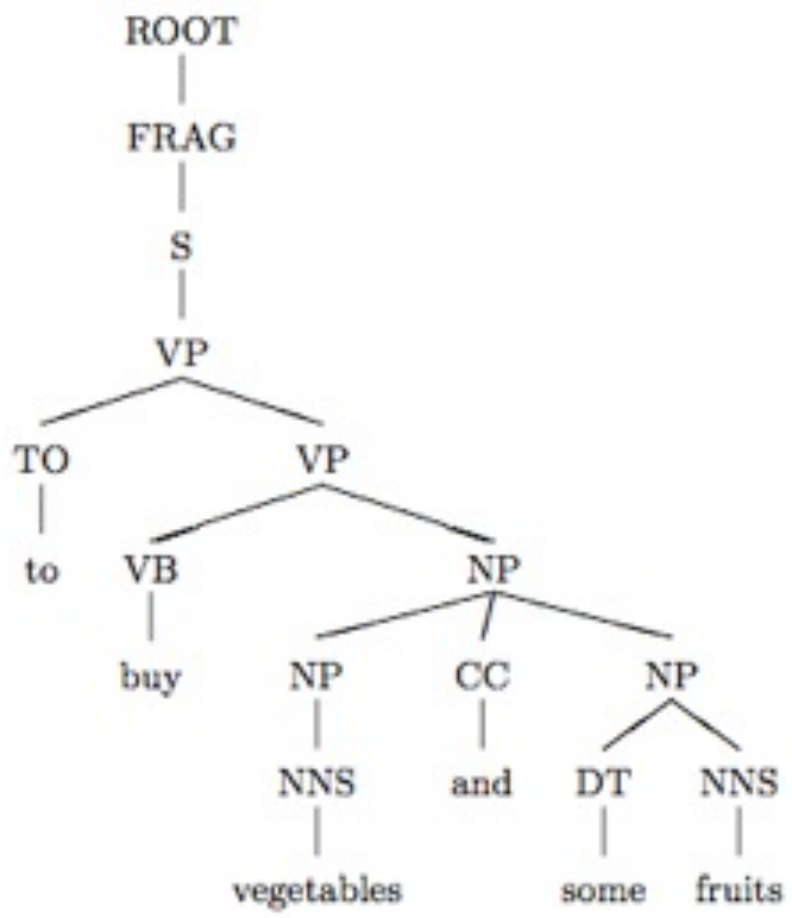




Chunking Text

I am going to the market to buy vegetables and some fruits





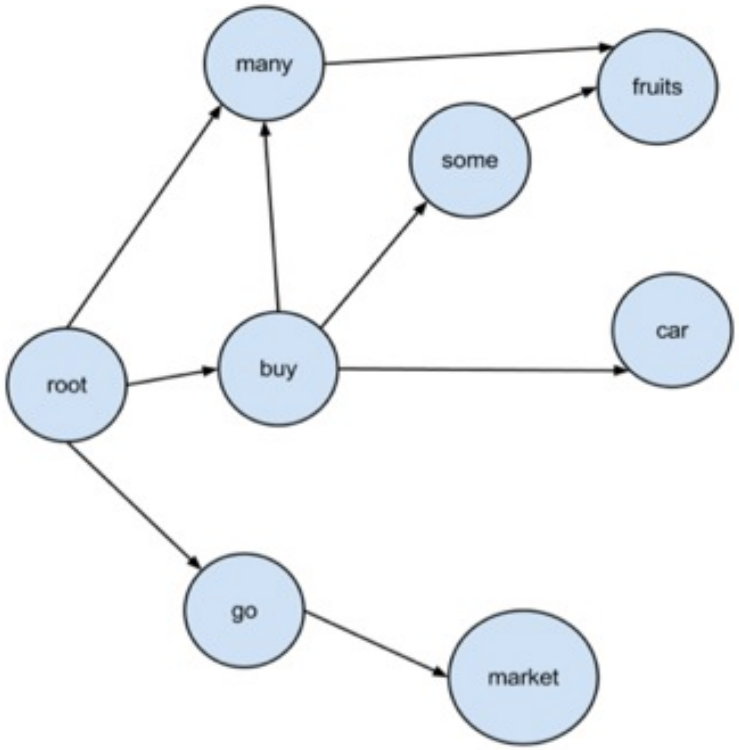
Candidate Spotting

After chunking and stemming, each potential noun chunk associated with individual verb chunks is paired with the stemmed verb in order to detect multi-word expressions of the form 'verb plus object'.

```
Data: NounPhrase
Result: Valid object concepts
Split the NounPhrase into bigrams ;
Initialize concepts to Null ;
for each NounPhrase do
    while For every bigram in the NounPhrase do
        POS Tag the Bigram ;
        if adj noun then
            add to Concepts: noun, adj+noun

        else if noun noun then
            add to Concepts: noun+noun

        else if stopword noun then
            add to Concepts: noun

        else if adj stopword then
            continue

        else if stopword adj then
            continue

        else
            Add to Concepts : entire bigram
        end
        repeat until no more bigrams left;
    end
end
```

Candidate Selection

In order to capture event concepts, matches between the object concepts and the normalized verb chunks are searched. This is done by exploiting a parse graph that maps all the multi-word expressions contained in the knowledge bases.



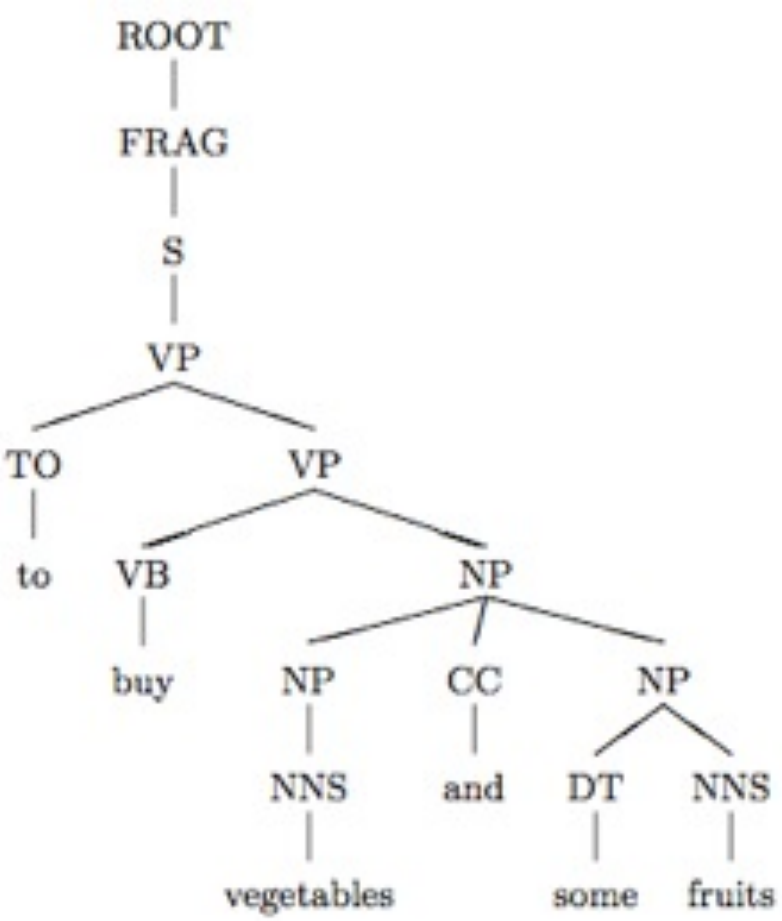
Concept Extraction

Candidate spotting

- *buy*
- *buy vegetable*
- *buy fruit*
- *vegetable and fruit*
- *buy vegetable and fruit*

Candidate selection

1. *buy vegetable and fruit*
2. *buy vegetable; buy fruit*
3. *buy; vegetable and fruit*
4. *buy; vegetable; fruit*



Similarity Detection

Because natural language concepts may be expressed in a multitude of forms, it is necessary to have a technique for defining the similarity of multi-word expressions so that a concept can be detected in all its different forms.

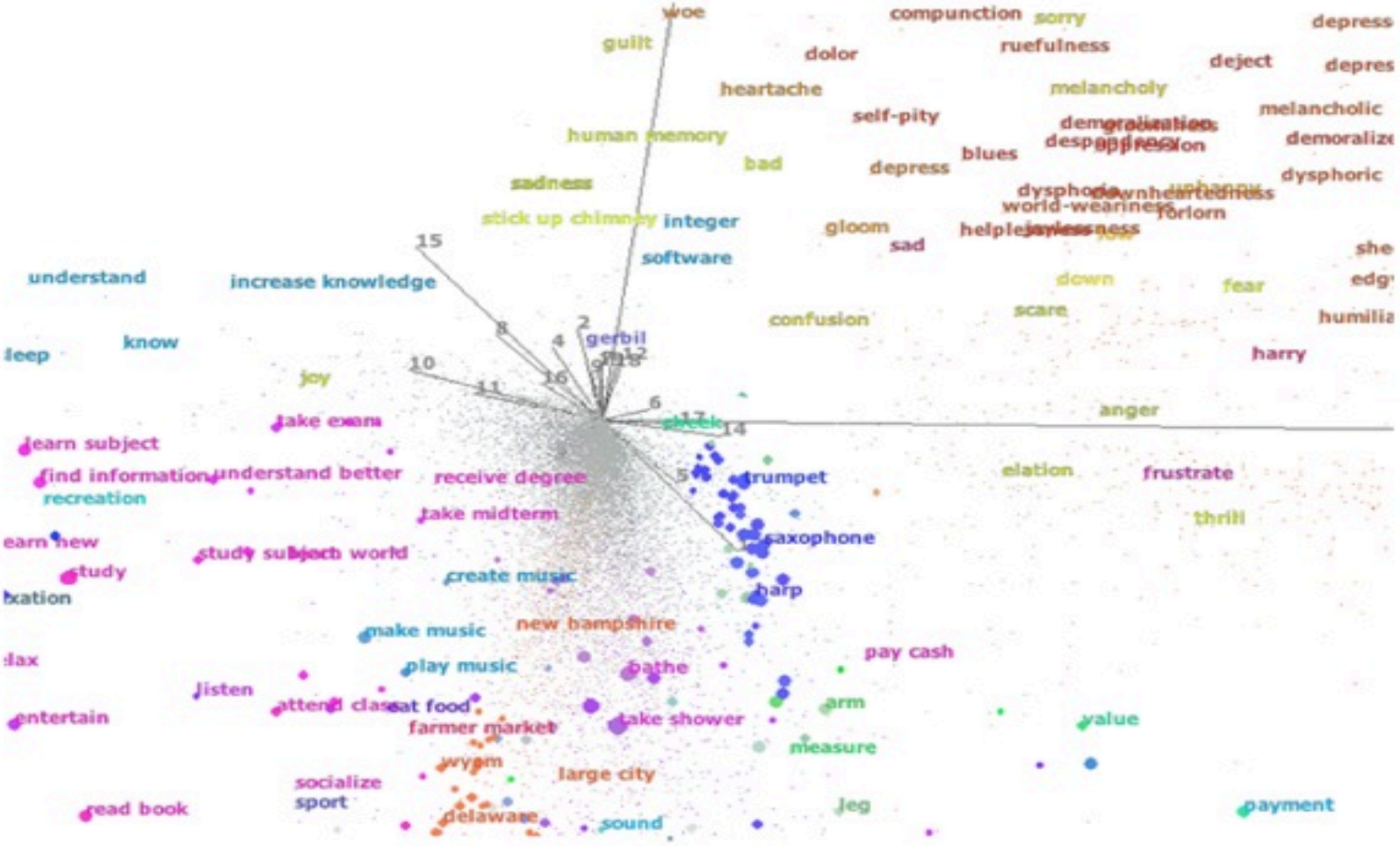
Data: NounPhrase1, NounPhrase2
Result: *True* if the concepts are similar, else *False*
if *Both phrases have atleast one* noun *in common* **then**
  Objects1 := All Valid Objects for NounPhrase1;
  Objects2 := All Valid Objects for NounPhrase2;
  M1 = matches from KB for
  M1 := ∅ ;
  M2 := ∅ ;
  for *all* concepts *in NounPhrase1* **do**
    M1 := M1 ∪ *all property matches for concept*;
  end
  for *all* concepts *in NounPhrase2* **do**
    M2 := M2 ∪ *all property matches for concept* ;
  end
  SetCommon = M1 ∪ M2 ;
  if *length of* SetCommon > *0* **then**
    The Noun Phrases are similar
  else
    They are not similar
  end

Semantic Similarity

Objects	Properties *(with simplified form)*					
	...	contains knowledge *contain knowledge*	has pages *have page*	is cold *be cold*	is for reading *be read*	...
⋮		⋮	⋮	⋮	⋮	
book	...	x	x		x	...
ice	...		-	x		...
newspaper	...	**x?**	**x**		**x**	...
magazine	...	x	x		x	...
⋮		⋮	⋮	⋮	⋮	

[5] E. Cambria, Y. Song, H. Wang, and N. Howard. Semantic multi-dimensional scaling for open-domain sentiment analysis. IEEE Intelligent Systems, doi:10.1109/MIS.2012.118 (2013)

AffectiveSpace



[4] E. Cambria and A. Hussain. Sentic Computing: Techniques, Tools, and Applications. Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2 (2012)

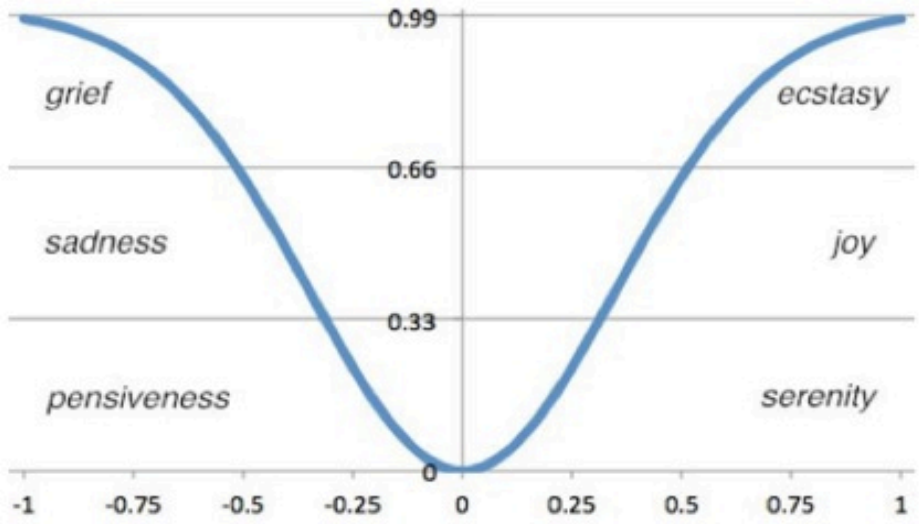
Hourglass Model

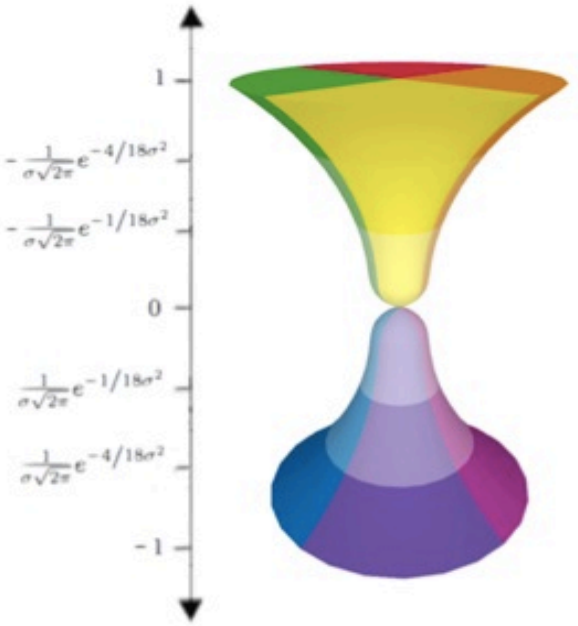
Human mind is made of different independent resources and emotional states result from turning some set of these on and turning another set of them off, changing how we think and see things.

[6] M. Minsky. The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. New York: Simon & Schuster (2006)

Hourglass Model

Interval	Pleasantness	Attention	Sensitivity	Aptitude
[G(1), G(2/3))	ecstasy	vigilance	rage	admiration
[G(2/3), G(1/3))	joy	anticipation	anger	trust
[G(1/3), G(0))	serenity	interest	annoyance	acceptance
(G(0), –G(1/3)]	pensiveness	distraction	apprehension	boredom
(–G(1/3), –G(2/3)]	sadness	surprise	fear	disgust
(–G(2/3), –G(1)]	grief	amazement	terror	loathing

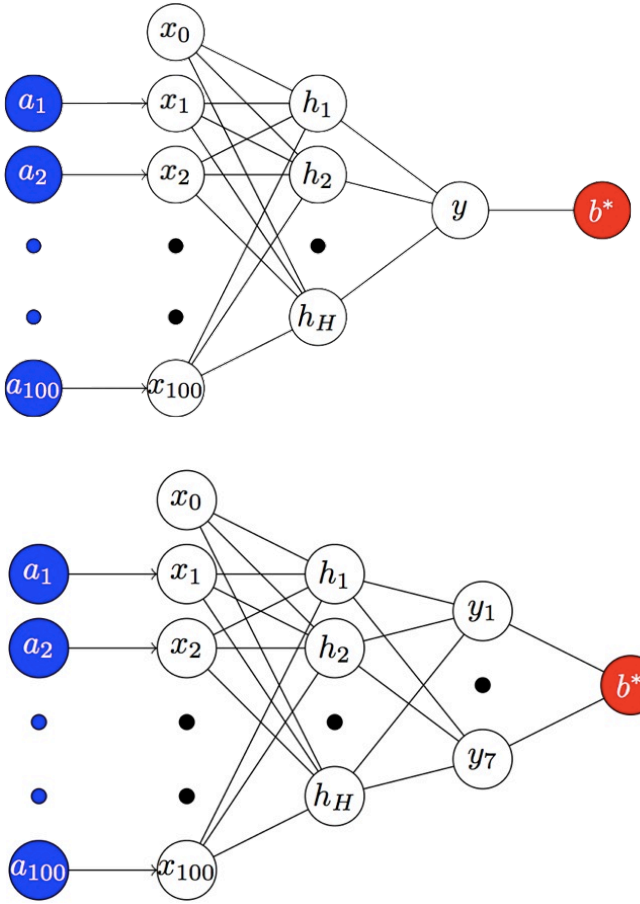




[7] E. Cambria, A. Livingstone, and A. Hussain. The hourglass of emotions. In: Cognitive Behavioral Systems, LNCS, vol. 7403, pp. 144-157, Springer (2012)

Sentic Neurons

The integration of a bio-inspired paradigm with PCA helps to better grasp the non-linearities of AffectiveSpace and, hence, improve the reasoning capabilities of the overall system.



[8] E. Cambria, T. Mazzocco, and A. Hussain. Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining. Biologically Inspired Cognitive Architectures 4, pp. 41-53 (2013)

Evaluation

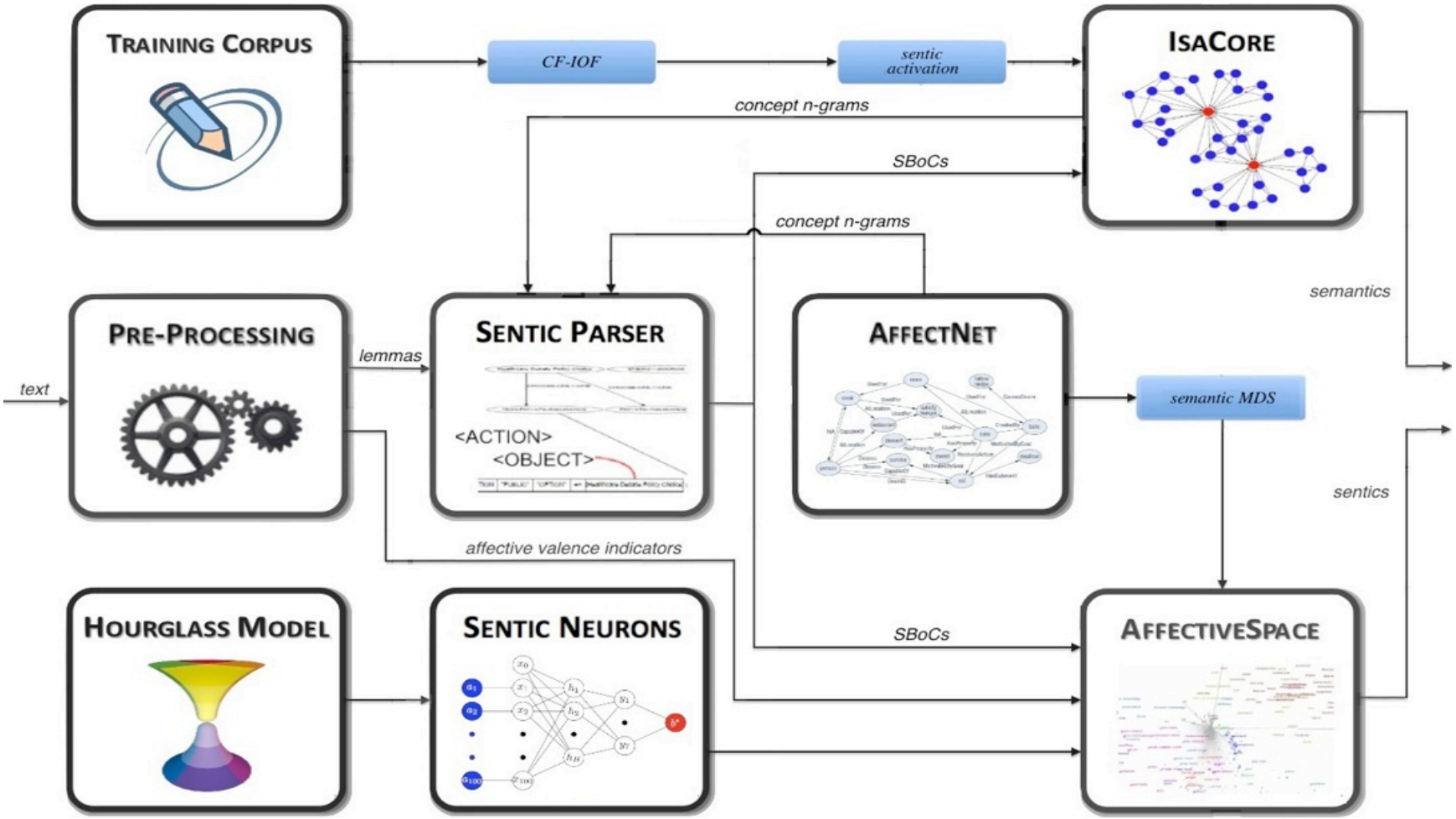
Algorithm	Precision	Recall	F-measure
Syntactic similarity	65.6%	67.3%	66.4%
Semantic similarity	77.2%	70.8%	73.9%
Ensemble similarity	85.4%	74.0%	79.3%

Table 1: Performance of different similarity detection algorithms over 200 concept pairs

Algorithm	Concept extraction accuracy
Naïve parser	65.8%
POS-based bigram	79.1%
POS-based + similarity	87.6%

Table 2: Performance of different parsing algorithms over 50 natural language sentences

Sentic Computing



[4] E. Cambria and A. Hussain. Sentic Computing: Techniques, Tools, and Applications. Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2 (2012)

Big Social Data Analysis

http://sentic.net/api

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about="http://sentic.net/api/en/concept/love">
        <rdf:type rdf:resource="http://sentic.net/api/concept"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/lust"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/love another person"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/sexuality"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/beloved"/>
        <semantics rdf:resource="http://sentic.net/api/en/concept/show empathy"/>
        <pleasantness rdf:datatype="http://www.w3.org/2001/XMLSchema#float">+0.491</pleasantness>
        <attention rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.0</attention>
        <sensitivity rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.0</sensitivity>
        <aptitude rdf:datatype="http://www.w3.org/2001/XMLSchema#float">+0.458</aptitude>
        <polarity rdf:datatype="http://www.w3.org/2001/XMLSchema#float">+0.316</polarity>
    </rdf:Description>
</rdf:RDF>
```

[9] E. Cambria, D. Rajagopal, D. Olsher, and D. Das. Big social data analysis. In: R. Akerkar (ed.) Big Data Computing, ch. 13, Taylor & Francis (2013)

Announcements

If you are interested in sentic computing,
please visit *http://sentic.net*

Feel free to play with the sentic API
http://sentic.net/api
and, if you find any bug, please tell me
cambria@media.mit.edu

A tutorial on sentic computing is going to be delivered tomorrow morning (Room Queluz VII)

Announcements

Please consider submitting to the Elsevier NNs special issue on Affective and Cognitive Learning Systems for Big Social Data Analysis

http://sentic.net/affcog

and to the Cognitive Computation special issue on Sentic Computing

http://sentic.net/cogcomp