

A Simple Strategy to Start Domain Ontology from Scratch

Ivo Wolff Gersberg / Nelson F. F. Ebecken

COPPE

Federal University of Rio de Janeiro
Rio de Janeiro, Brazil

Abstract—Aiming the usage of Domain Ontology as an educational tool for neophyte students and focusing in a fast and easy way to start Domain Ontology from scratch, the semantics are set aside to identify contexts of concepts (terms) to build the ontology. Text Mining, Link Analysis and Graph Analysis create an abstract rough sketch of interactions between terms. This first rough sketch is presented to the expert providing insights into and inspires him to inform or communicate knowledge, through assertive sentences. Those assertive sentences subsidize the creation of the ontology. A web prototype tool to visualize the ontology and retrieve book contents is also presented.

Keywords— domain ontology; contextual approach; ontology; NLP

I. INTRODUCTION

Since 1990s Domain Ontology was seen as a way to formally model a system's structure. An ontology engineer seeks to represent specific knowledge, analyse the most relevant entities (more general and abstract entities that can be subdivided into categories such as objects, processes, and ideas) and organizes them into concepts and relationships. The skeleton of ontology consists of a hierarchy of generalized and specialized concepts [1].

Undeniably, different currents are found on the discovery of textual patterns for building ontology. On the one side are the experts of Natural Language Processing (NLP) and reasoning, claiming that a semantic approach is mandatory for dealing with ontologies. One of challenges during the transformation from data to knowledge is the use of semantic instead of traditional Text Mining techniques [2].

On the other hand, those that advocates the use of simple Text Mining statistical techniques. NLP could be used for an understanding of the analysis or synthesis of texts and not necessarily for an understanding of the texts [3]. Therefore, the Information Retrieval area and the NLP began to share algorithms and statistical methods with the help of lexical dictionaries to provide answers for relatively elaborate subjects. Many scientists still contend that these statistical methods are inadequate for contextual knowledge extraction; however, for certain purposes, they are reasonably efficient [4].

It is clear that the line is quite blurry thought between a genuine NLP tool and a statistical tool. How to build domain ontology in a fast and easy way without the use of

sophisticated semantic tools and that it serves from aid to the study of neophytes of a certain subject is the main objective of this work.

One of the significant challenges facing an ontology engineer is that, in most instances, he does not possess the specific domain knowledge that is the subject. He does not know which concepts in a scientific text are important and how to start talking with an expert. By overcoming this shortcoming, the proposed methodology intends to provide a preliminary and abstract rough sketch of interactions between terms obtained from unstructured data (domain classical books), using automatic techniques of Text Mining, Link Analysis and Graph Analysis, based on contexts. This first rough sketch is presented to the expert providing insights into and inspires him to inform or communicate knowledge to the ontology engineer and construct Domain Ontology using two hands.

Once the ontology building was based in a middle-out strategy [5], in which concepts were generalized and specialized, and ontologies are always a work in progress, the main task is identify concepts (terms) that can relate to provide short assertive sentences (identified by the expert). The ontology engineer alone would not be able to identify and build such sentences, but on terms that seem obvious to expert's eyes, the task becomes easier. How to get these concepts, i.e., how to extract terms from a non-structured data and how to presents those in a suggestive way to an expert, is the goal to build Domain Ontology from scratch.

II. BUILDING DOMAIN ONTOLOGY METHODOLOGY

The proposed methodology, focused on practicality, utilizes two different tools: *PolyAnalyst Data Analysis* (PA)¹ for Information Retrieval/Link Analysis and *Gephi Graph Visualization*² for Graph Analysis.

A. Corpuses and extracting concepts

The first task determined that the Domain Ontology will be created about a mathematical subarea: Fractal. The resources for obtaining the relevant concepts were chosen by the expert and were composed of nine classical books on mathematical fractal, with an average of 340 pages and a total of 680,000 words (after pre-processing). It was considered using all chapter contents of the books and also

¹ Megaputer Intelligence, Inc.

² Gephi Consortium, CDDL and GNU General Public License

using only the words within the indexes of the same books (as concepts suggested by the authors - virtual specialists). Two corpuses were built: the FRACTAL Corpus (148 documents originated from the 148 chapters of the adopted books) and the Index Corpus (nine documents originated from the indexes of the adopted books). The Text Mining techniques were applied to these two corpuses separately.

The task of *Concept Acquisition* and *Selection* reveals the terms (nouns) considered essential to fractal knowledge. The strategy for this phase involved approaches without an expert's presence, starting from a set of terms (unigram and bigram) originated from PA tool. Those terms were measured by a significance value to represent how different a word is in all the texts, a measure 'above' as an average from the simple word frequencies, which is compatible with the classical measure Tf-IDF — term frequency–inverse document frequency, an excellent example of a statistical index that gives quantitative answer as to whether a term is really worth being extracted [6]. The sets were normalized, ranked and pruned by a high threshold.

B. Contexts to build assertive sentences

Contexts are abstract objects and difficult to be defined [7], normally examples are offered, but every communication needs contexts because without contexts there is no meaning [8]. Let's consider an example: give the words *generator* and the word *tree* to a person without fractal knowledge, so this person cannot think towards a fractal context. Probably he will say, "Ok, a machine that converts one form of energy into another using trees", not ecological and perhaps not an assertive sentence. But a person thinking inside fractal knowledge, immediately and without effort will say "a generator that has a line segment, as an initiator, will construct a fractal tree".

Various notions will be found about what context is and how to treat it formally. Ramanathan Guha, over McCarthy works, tried to give a concept of context: Contexts are objects in the domain, i.e., we can make statements about contexts, as in [9]. A specific formula to treat contexts is used by McCarthy and Guha, based on sentences of the form: $ist(c,p)$, where ist stands for 'is true in' and is to be taken as assertions that the proposition p is true in the context c . Considering the idea of construct assertive sentences, but not operations using Reasoning and First Order Language as McCarthy and Guha, our question is *how to construct assertive sentences?* The unstructured data of the books creates our universe of discourse and this data in a structure of a graph is presented to the expert, aiming for immediate recognition of meanings in a short assertive sentence. These short sentences will point the concepts and relations, of the ontology (taxonomic or non-taxonomic).

The first approach to construct the contexts used the Link Analysis technique from the PA tool. Using the sets of terms, generated in the previous section, as single-level taxonomy, we applied them to the corpuses to obtain representative labels for each of its documents. In this way, the documents were deconstructed to a few isolated terms

(concepts). Similar to reduced labels documents, favouring a computational gain, these words were submitted to analysis using the Link Analysis technique, looking for correlation patterns for which the connections among the vertices of the graphs are measured by tension values. The undirected graph generated was shown to the expert, who manually identified possible contexts, like cluster of words with something in common. In a second moment, an automatic method using Graph Analysis technique, Community Detection, was used to identify possible contexts and to compare the expert results. Community Detection is more often used as a tool for the analysis and understanding the structure of a network, for shedding light on patterns of connection that may not be easily visible in the raw network topology, as in [10]. Gephi tool offers the possibility of mixing two different techniques: Community Detection and Laplacian Dynamics. The first one requires partitioning a network into communities of densely connected nodes, with the nodes belonging to different communities that are only sparsely connected. The quality of the partitions resulting from these methods is often measured using the so called modularity of the partition [11]. The second one introduce the stability of a network partition as a measure of its quality in terms of the statistical properties of a dynamic process occurring on the graph, instead of the structural properties [12]. The tension values of the Link Analysis were used to weigh the network links. To be fair, with the number of communities manually marked by the expert, this number was controlled by the resolution factor parameter given by the Laplacian dynamics method. Once in possession of contexts, the expert identified and constructed assertive sentences about fractal knowledge. These assertive sentences help the ontology engineer to build the ontology hierarchisation.

III. ONTOLOGY VISUALISATION

The construction and use of generic ontologies provides an alternative method for searching for and visualising a desired portion of knowledge. Instead of search concepts and documents based only on keywords, it is possible to search by *context*. A Web search engine prototype was created, allowing visualisation of the *contexts* of the ontology and document (chapter) retrievals, providing fractal neophytes with an initial path for their studies. The implementation, through the Thinkmap³ tool based on Graph Theory, allows for a visualisation of the created ontology as an oriented graph among the concepts (vertices) and the relationships (edges). The document retrieval is based on relationships, revealing the FRACTAL Corpus's most relevant chapters through the well-known algorithm Vector Space Model (VSM).

IV. RESULTS

The task of *Concept Acquisition* and *Selection* was automatically applied over the two corpuses separately, limited by the high threshold, giving two term sets. The expert also manually chose concepts over an unranked set, generated from the contents of the books, without any kind of pruning, only to check the performance of the automatic

³ Thinkmap Visualize Complex Information — Thinkmap, Inc.

extraction. Considering the expert's choices and only using the indexes of the books, we obtained 48% and 23% of terms in common for unigrams and bigrams, respectively, i.e., regular results. However, if we think that the indexes of the books were only words, totally unstructured, without any sentence and very short file size, probably we can use them alone in some occasions. Using only the contents of the books, it was obtained 100% and 32% of terms in common for unigrams and bigrams, respectively. It was decided to aggregate the contents books set and the terms that was in the indexes set and not in the contents of the books set, obtaining 100% and 47% of terms in common with the expert's choices for unigrams and bigrams, respectively. Therefore, the strategy for joining those terms of the book indexes and the term set of the contents had better performance; in other words, this strategy improved the results compared with only using Text Mining of the contents or only the scenario that included the indexes of the books.

Once in possession of the unigram and bigram terms as the final set of concepts (590 terms), the task of *context detection* was performed by Link Analysis and Detection Community.

A. Building contexts and the Fractal Domain Ontology

Over the final set of concepts, the Link Analysis technique generated an undirected graph. This graph was presented to the expert and he outlined, by hand, possible contexts of the big fractal context (blue colour in **Error! Reference source not found.**). Those contexts gave the idea of the most generic concepts of the Fractal Ontology and the concepts inside the contexts were used to construct the assertive sentences. Aiming an automatic process to identify the contexts, the Community Detection of Networks was applied over the Link Analysis results (Fig. 2). Numerous similarities were observed between the contexts manually signed by the expert in the graph of the Link Analysis and the contexts as communities in the network of the Graph Analysis (0). The contexts (communities) automatically detected was presented again to the expert. The collaborative job between the expert and the ontology engineer to construct the ontology is shown in TABLE III. , with few examples per context, but of course numerous other assertive sentences were created.

B. Comparison with a semantic space model

The results were also compared with a semantic technique. The BEAGLE Model was applied using a word similarity visualization tool, Word2Word⁴, where words are represented by high-dimensional holographic vectors. An environmental vector is created to represent the physical characteristics of words in the environment (e.g., orthography, phonology, etc.), whereas the memory vector represent internal memory for co-occurrence and position relative to other words using convolution and superposition mechanisms [13]. In this case, we used all the words of our universe of discourse ($\approx 680,000$

words), showing concepts in a distribution based in a similarity distance.

The intention was observe if one concept inside a context has his neighbours closer together in a semantic similarity distribution, i.e., closer together with higher similarity metric.

Based in the semantic space created, TABLE I. shows the top neighbours to fractal, power law, probability density and iteration (concepts extracted from different contexts found in Fig. 2). In this space we have all kind of words like nouns, verbs, *stoplist* words, etc., but if we check only for nouns, we can observe that some of them (highlighted) are also in his respectively context found. The highlighted common terms found are, indeed, the most important concepts pointed by the expert viewpoint to construct the assertive sentences.

TABLE I. TOP NEIGHBORS IN THE SEMANTIC SPACE

fractal	power law	probability density	iteration
distance to	distance to	distance to	distance to
0.619 <i>selfsimilar</i>	0.378 fractal	0.360 <i>autocorrelation</i>	0.566 fractals
0.601 <i>fractal-dimension</i>	0.326 <i>scaling</i>	0.309 graph	0.561 <i>equation</i>
0.578 set	0.324 fractaldimension	0.294 nondecreasing	0.551 <i>juliaset</i>
0.574 equation	0.322 probability	0.293 weierstrass	0.545 fractaldimension
0.567 probability	0.309 important	0.282 <i>brownian</i>	0.533 set
0.561 fixedpoint	0.302 integer	0.279 this	0.526 point
0.561 point	0.302 point	0.267 reason	0.526 <i>fixedpoint</i>
0.552 chaotic	0.300 simple	0.263 fractal	0.525 <i>cantorset</i>
0.545 important	0.293 proportional	0.262 convex	0.524 example
0.549 <i>measure</i>	0.293 set	0.257 cases	0.523 result
0.542 <i>fractals</i>	0.292 selfsimilar	0.256 result	0.522 follows
0.539 general	0.290 equation	0.253 note	0.520 cases
0.538 cases	0.289 equivalent	0.252 general	0.517 plate
0.535 follows	0.288 obtained	0.250 fractal-dimension	0.517 system
0.535 simple	0.288 problem	0.248 juliaset	0.510 sierpinski-gasket
0.532 example	0.288 follows	0.247 <i>probability</i>	0.507 dimensions
0.531 result	0.284 chaotic	0.245 suppose	0.504 <i>attractor</i>
0.527 juliaset	0.283 constant	0.245 similarly	0.502 however
0.524 <i>selfsimilarity</i>	0.282 similar	0.245 write	0.494 on

Another way to see the semantic space is laying out the nodes using Multidimensional Scaling (MDS) algorithm according to similarity relationships (Fig. 5). For example, observing the 100 near similarity terms in the semantic space for concept *iteration*, it was found some important concepts (red nodes) that were also found in the *context approach*.

Others nouns (concepts) that are around the *iteration* concept were found in different contexts, but this fact is not a big problem because the goal is to construct Domain Ontology in a middle-out strategy. This suggest us (for future works) a way to link the contexts of our approach, i.e., construct assertive sentences using a concept from one context and another from other context.

⁴ Word2Word (W2W) — Kievit-Kylar,B., Cognitive Computing Lab, Indiana University, USA

TABLE II. MANUAL X AUTOMATIC IDENTIFICATION OF CONTEXTS

manually contexts concepts		automatic contexts concepts	
upper right corner	attractor, nonlinear, chaos, chaoticbehavior , bifurcation, lorenzattractor, deterministicfractal sensitivedependence,	red context	attractor, nonlinear, chaos, chaoticbehavior , bifurcation, lorenzattractor, deterministicfractal, sensitivedependence,
upper middle	iteratedfunctionssystem, iteration, complexplane, mandelbrotset, juliaset	red context (middle)	iteratedfunctionssystem, iteration, complexplane, mandelbrotset, juliaset
upper left corner	borelset, boxcountingdimension, packingdimension, boxdimension	red context (upper)	borelset, boxcountingdimension, packingdimension, boxdimension
lower left corner	generator, initiator, curve, cantorbar, cantordust, snowflake, triadickoch, kochisland, triadic, midpoint, peano	pink context (upper)	generator, initiator, curve, cantorbar, cantordust, snowflake, triadickoch, kochisland, triadic, midpoint, peano
lower right corner	sierpinski-gasket, dragoncurve brownianmotion, fractionalbrownian, randomwalk, randomfractal, probabilitytheory, distributionnormal, whitenoise	yellow context	sierpinski-gasket, dragoncurve brownianmotion, fractionalbrownian, randomwalk, randomfractal, probabilitytheory, distributionnormal, whitenoise
right middle	powerlaw, exponent, scalinglaw, distribution, frequency	magenta context	powerlaw, exponent, scalinglaw, distribution, frequency, logarithm
middle	fractal, fractalgeometry, fractalset, mandelbrot, dimension, dimensiond, selfsimilarity, coastline, topologicaldimension, hausdorffdimension	pink context	fractal, fractalgeometry, fractalset, mandelbrot, dimension, dimensiond, selfsimilarity, coastline, topologicaldimension, hausdorffbesicovitchdimension

TABLE III. FEW EXAMPLES OF ASSERTIVE SENTENCES CREATED BY THE EXPERT

context	assertive sentence	ontology concepts	relation
pink	All generators are obtained by an initiator.	Generator –Initiator	hasInitiator
magenta	Self-Similarity is one of the most important property of fractal objects.	Fractal – SelfSimilarity	hasProperty
	Fractional Brownian motion is a random walk process.	FractionalBrownianMotion – BrownianMotion	is_a
cyan	Fractal Dimension is calculated by a Power Law function.	BrownianMotion – RandomWalk	is_a
	Power Law has a mathematical property that obeys a Scaling Law, scale invariance.	DimensionD – PowerLaw	isCalculatedBy
red	The Mandelbrot Set and the Julia Set are set of points in the complex plane, created by iteration process. Iterated Function System (IFS). Rotation is a Linear Transformation.	PowerLaw – Function	is_a
		ScalingLaw – MathematicalProperty	is_a
		ScaleInvariance – ScalingLaw	is_a
		PowerLaw – ScaleInvariance	hasProperty
green	A Fractal can be a self-affine fractal or self-similar fractal; it means it can have an affine or a similar symmetry.	MandelBrotSet – SetOfPoint	is_a
		JuliaSet – ComplexPlane	isLocated
		Iteration – Process	is_a
		IFS – Iteration	is_a
		MandelbrotSet – IFS	isCreatedBy
		JuliaSet – IFS	isCreatedBy
		SelfAffineFractal – SelfAffine	hasSymmetry
		SelfSimilarFractal - SelfSimilar	hasSymmetry
		SelfAffine – AffineTransformation	hasTransformation
		SelfSimilar –LinearTransformation	hasTransformation

C. Ontology Visualisation Web Prototype

Once in the possession of the ontology, a prototype of searches oriented by contexts was implemented as a Web app, Fig. 3. The FRACTAL Domain Ontology visualization was created as an oriented graph among the essential concepts and the relations that clarify the fractal knowledge. The relationships of the taxonomic type are represented by grey edges, whereas the non-taxonomic relationships, which are knowledge in itself, are represented by pink colour. When passing the mouse on the relations, the tool indicates the specific name of the relation.

Fig. 3 enhanced the non-taxonomic relation *isCalculatedBy*, indicating that a fractal can be calculated by a power law. Clicking on a concept or looking for a certain concept (in the search area), leads to the tool unfolding new concepts related to the selected or searched for concept, presenting a context of relations appropriate to a certain desired granularity. The interface is easy to use, does not demand any previous knowledge, and is handled by clicking on concepts or on relations or by dragging on concepts. A sole concept does not clarify knowledge; thus, for the tool developed, only clicking on relations restores the chapters of the books.

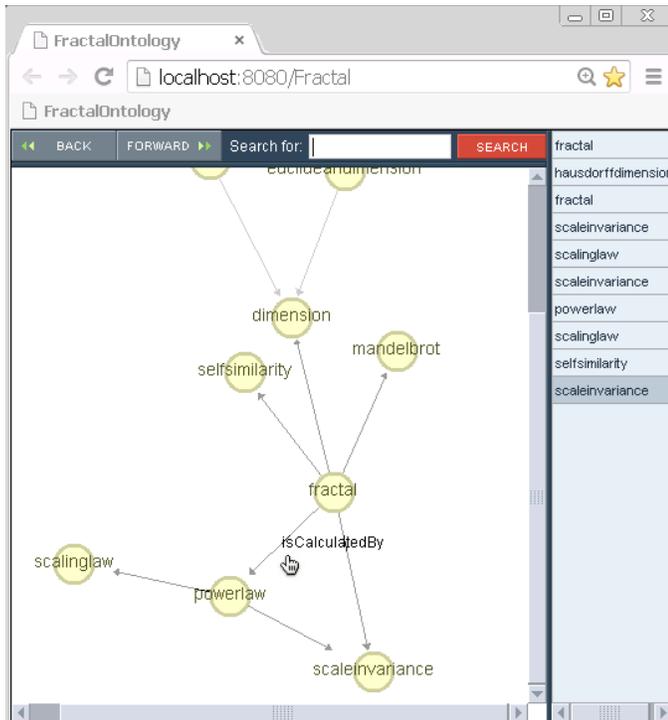


Fig. 3. A piece of Fractal Ontology Web Visualisation

When clicking on the enhanced relation, the tool presents an indication of bibliographical references for studies. The tool only returns the authors and the chapters of the books most relevant for associating the investigated relation, whereas the percentage relevance of each chapter is marked beside the selected chapter. The relevance is calculated as a modification of the Vector Space Model (VSM) technique, in which the vectors of the documents with a very small distance are considered similar to the relation. The more similar it is, the more relevant it will be in the document.

We now emphasise the notion of a contextual search in the following way: a student was interested in knowing where to learn how to calculate fractal dimensions through power laws. Using the example above (Fig. 3), a neophyte's glance into fractal knowledge by solely looking at the graph and moving the mouse onto the concepts, will see that fractals are calculated using a power law. Clicking on the relation *isCalculatedBy*, the chapters 1, 4, 5 and 17 of the book *Fractal, Chaos and Power Laws* was obtained as a result, whose author is Schroeder. The tool restricts the number of documents returned to those with a relevance of 99.5% or higher. For a student who is not curious and not stimulated by contextual reasoning, the task is concluded and he/she only studies those book or chapters. However, by consulting an expert, we found that among our nine books that have the most suitable chapters for understanding the calculus of fractal dimensions by power laws, he advised the following: chapter 2 of the book *Fractals* (author Feder) and chapter 4 of the book *Fractal, Chaos and Power Laws* (author Schroeder).

The usage of ontologies based on a search oriented by contexts will amplify the understanding of the subject. A student should be motivated to analyse the context in which the concepts *fractal* and *power law* are involved, as shown in

Fig. 3. It is shown that the concept *power law* was verified to have a relationship with the concept *scale invariance* and with the concept *scaling law*; yet, the concept *fractal* was also related to the concept *scale invariance*. Within this context, we anticipate that the ontology relationship among the concepts *power law* and *scale invariance* is *hasProperty*. Therefore, if the student notices that the relation *hasProperty* is more intrinsic than *isCalculatedBy*, he/she will be urged to relationally analyse the concepts *power law* and *scale invariance*. Selecting this last relation, he/she will obtain Fig. 4 as a result.

Thus, the student notices that the most relevant chapters of this relation are those with a relevance of higher than 99.9%; among these are chapter 2 of Feder's book and chapter 4 of Schroeder's book, which the expert verbally recommended. In the event that the student chose these books or chapters for study, such choices would have agreed with the expert's indications without having consulted him. Once there are no problems with copyrights for certain books, clicking the desired chapter/book enables it to be consulted digitally.

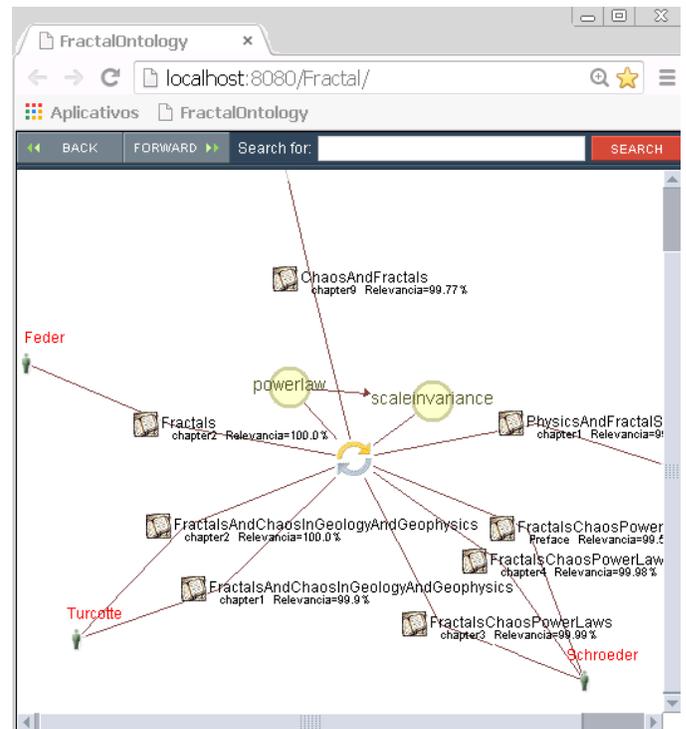


Fig. 4. powerlaw-scaleinvariance relationship (relevant chapters)

V. FINAL CONSIDERATIONS

The usage of ontology for us is not dedicated to reasoning, Q&A or merging databases. The main purpose, adopted in this work, was to build an educational Domain Ontology from scratch of the chosen subject in a fast and easy way. Based in classical books and having a vast consensual bibliography about the subject in question, the neophytes (students) can have an efficient retrieval method through a visual ontology web tool. In order to share knowledge, textual evidence needs to be linked to ontologies as the main repositories of represented knowledge [14].

REFERENCES

A simple statistical approach looks for terms in a statistical way inside a text, while semantic information looks for terms whose grammar and syntax rules reveal some semantics. A semantic information need to know at least the phrase that contains the term in question, once words are characterized by the company it keeps [6]. Depending on the windows size of words to look at, the semantic approach will be time consuming.

The methodology used in this work gave attention to identify important terms together giving contextual meaning, using only simple statistical techniques (Tf-IDF and correlations of terms as a Link Analysis graph) and a network (communities) representation.

Emphasizing that the present methodology is a semi-automatic approach and accordingly to the fractal expert, the *contexts* created revealed us well an enough concepts to start a new ontology. A semantic distribution, like BEAGLE model, not offered great advantages in the present case.

Therefore, a simple way using a classical term extraction, Link Analysis and Community Detection can be used to start Domain Ontology from scratch, using only classical books about the subject in question.

ACKNOWLEDGMENT

We would like to give special thanks to Emeritus Professor Luiz Bevilacqua (COPPE/Federal University of Rio de Janeiro), who helped us through his notable knowledge of fractals. The authors acknowledge the support provided by CNPq, the Brazilian Research Agency, and FAPERJ, the Rio de Janeiro Research Foundation.

- [1] N. Guarino, D. Oberle, and S. Staab, "What Is an Ontology?," in *Handbook on Ontologies*, Springer-Verlag, 2009.
- [2] H. Jingshan, D. Dou, J. Dang, J. H. Pardue, X. Qin, J. Huan, W. T. Gerthoffer, and M. Tan, "Knowledge acquisition, semantic text mining, and security risks in health and biomedical informatics," *World J. Biological Chem.*, vol. 3, no. 2, pp. 27–33, Feb. 2012.
- [3] M. Konchady, *Text Mining Application Programming*. Massachusetts: Charles River Media, Thomson Learning Inc., 2006.
- [4] Z. Li, M. C. Yang, and K. Ramani, "A methodology for engineering ontology acquisition and validation," *Artif. Intell. Fo Reng. Des. Anal. Manuf.*, no. 23, pp. 37–51, 2009.
- [5] M. Uschold and M. King, "Towards a Methodology for Building Ontologies," presented at the IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.
- [6] A. Stavrianou, P. Andritsos, and N. Nicoloyannis, "Overview and Semantic Issues of Text Mining," *SIGMOD Rec.*, vol. 36, no. 3, pp. 23–34, Sep. 2007.
- [7] J. McCarthy, "notes on formalizing context," 13-Apr-2005.
- [8] G. Bateson, *Mind and Nature: A Necessary Unity*. New York: Dutton, 1979.
- [9] R. V. Guha, "Contexts: a formalization and some applications," Stanford University, 1995.
- [10] M. Newman, *Networks: An Introduction*, 1st ed. Oxford University Press, USA, 2010.
- [11] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *arXiv:0803.0476*, Mar. 2008.
- [12] R. Lambiotte, J.-C. Delvenne, and M. Barahona, "Laplacian Dynamics and Multiscale Modular Structure in Networks," *arXiv:0812.1770*, Dec. 2008.
- [13] M. N. Jones and D. J. K. Mewhort, "Representing word meaning and order information in a composite holographic lexicon.," *Psychol. Rev.*, vol. 114, no. 1, pp. 1–37, 2007.
- [14] I. Spasic, S. Ananiadou, J. McNaught, and A. Kumar, "Text mining and ontologies in biomedicine: Making sense of raw text," *Brief. Bioinform.*, vol. 6, no. 3, pp. 239–251, Sep. 2005.

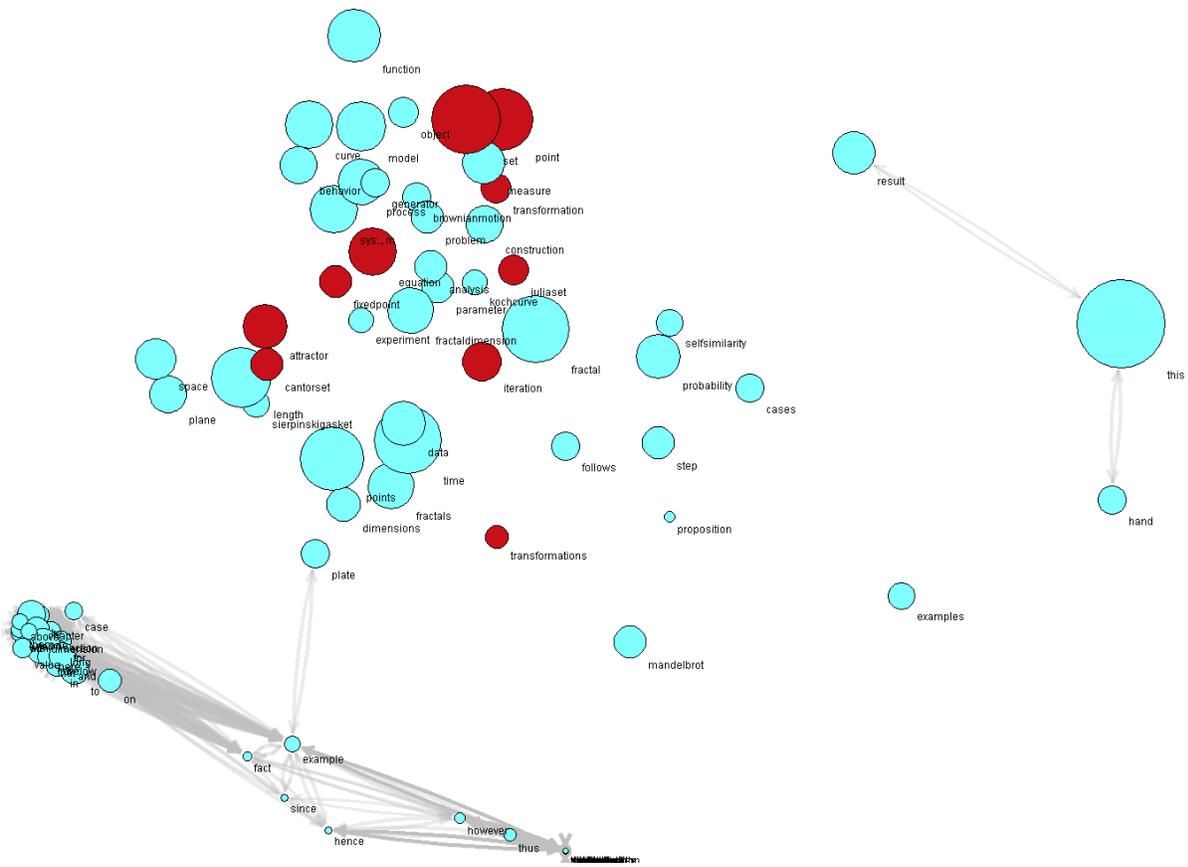


Fig. 5. Semantic Distribution of Concept *iteration*