

# THE INDUSTRIAL INNOVATION PROCESS BY A NETWORK APPROACH

Nelson F. F. Ebecken<sup>(a)</sup>, Francisco A. Moreira Jr.<sup>(b)</sup>

<sup>(a)</sup> COPPE/Federal University of Rio de Janeiro

<sup>(b)</sup> COPPE/Federal University of Rio de Janeiro

<sup>(a)</sup>[nelson@ntt.ufrj.br](mailto:nelson@ntt.ufrj.br), <sup>(b)</sup>[fmoreira@gmail.com](mailto:fmoreira@gmail.com)

## ABSTRACT

The objective of this paper is to present the methodology adopted by Petrobras, the Brazilian oil company, in e-science collaborative projects to generate ideas of new opportunities obtained by technology convergence for an energy company that searches solutions for the future. The effort is based on contributions of a group of 50 senior scientists from top 5 research institutes working iteratively via the Web and text mining activities to look for new technologies.

Keywords: text mining, collaborative project, convergent technologies

## 1. INTRODUCTION

The convergence of nanotechnology, biotechnology, information technology, and cognitive science is transforming global society. Technological convergence is beginning to define the way societies interact themselves, the way science is done and the way the global marketplace is organized.

In middle 2006, Petrobras, a Brazilian oil company, established a pioneer multidisciplinary team of senior academics and end users working in partnership to find potential opportunities and benefits of technology convergence.

A scientific advisory board was formed consisting of representatives from top universities as well as Petrobras. The present membership includes major government-funded Brazilian research institutes: Federal University of Rio de Janeiro, Federal University of Rio Grande do Sul, State University of São Paulo, State University of Campinas and the Aeronautic Technology Institute. The function of the advisory board is to channel information from Petrobras to faculty members concerning problems that need research attention and to maintain the focus.

In terms of the technology to enable this kind of online scientific collaboration there is much to choose from: blogs, mailing lists, wikis, forums, consensus Web filters, etc. What is shocking are the cultural and organizational obstacles — not monetary or financial obstacles — to getting individuals and organizations working together.

Many academics have said online collaboration won't work because, to put it in a nutshell, a) publish or perish

means that work with no publications equals no motivation and b) who has time for this even if there was motivation? Also most of the scientists are pretty techno-adverse and they view hanging around on science blogs as a time sink.

In fact, most established PIs today were educated in very different circumstances to what undergraduates and postgraduates are today. From our observations most people will use the tools they are familiar with and find change very difficult. Many are unwilling to experiment with computers.

We've had so much push back from our non-geek colleagues on the blog and wiki idea that we started developing something different — a Web-based innovation engine which builds on respect for expertise, the scientific method, evidence-based texts, and, above all, the value of the individual scholar, author, and creator of knowledge.

## 2. IMPORTANT REMARKS

The start point to design this online collaboration tool was a relatively new breed of website that collaboratively votes on the best links, like Digg, Reddit, NowPublic and NewVines. Each of these popular aggregator sites supplies anonymous readers with pointers to news items that are ranked by other anonymous readers. None of these sites generates news; they only point to it by filtering the links to news items. Readers themselves vote on the importance of linked items suggested by other readers; these votes are then subjected to a complex formula to produce rankings.

That is part of the larger pattern of the appeal of a new online collectivism that is nothing less than a resurgence of the idea that the collective is all-wise. The so-called Digital Maoism is an unholy brew made up of digital utopianism; a desire to avoid individual responsibility; anti-intellectualism — the common disdain for pointy headed professors; and the corporatist "team" mentality (Gorman 2007, Lanier 2006).

The flight from expertise is accompanied by the opposite of expertise — the phenomenon that Andrew Keen has called, in his book of the same name, "the cult of the amateur," a world in which everyone is an expert in a world devoid of expertise.

The argument is that scholarship based on individual expertise resulting in authoritative statements is

somehow passé and that today's younger people think and act differently and prefer collective to individual sources because of their immersion in a digital culture. If we deviate from the party line — radical epistemic egalitarianism — then we are Elitists, pure and simple, and therefore the enemy.

It's safer to be the aggregator of the collective. You get to include all sorts of material without committing to anything. You can be superficially interesting without having to worry about the possibility of being wrong, except when intelligent thought really matters. In that case the average idea can be quite wrong, and only the best ideas have lasting value. Science is like that.

The structures of scholarship and learning are based on respect for individuality and the authentic expression of individual personalities. The person who creates knowledge or literature matters as much as the knowledge or the literature itself. The manner in which that individual expresses knowledge matters too.

Scientific communities achieve quality through a cooperative process that includes checks and balances, and ultimately rests on a foundation of goodwill and "blind" elitism — blind in the sense that ideally anyone can gain entry, but only on the basis of a meritocracy.

Our online collaboration tool was designed to support the idea that individual scholars matter, not just the process or the collective. The collective is more likely to be smart when it isn't defining its own questions, when the goodness of an answer can be evaluated by a simple result, and when the information system which informs the collective is filtered by a quality control mechanism that relies on individuals to a high degree. Under those circumstances, a collective can be smarter than a person. Break any one of those conditions and the collective becomes unreliable or worse.

### 3. INNOVATION ENGINE

The vast majority of idea generation methods in existence are a product of Worldview 1 thinking. Worldview 1 adherents think that if you want to have a brilliant idea, you must produce a large number of ideas and the brilliant one will be in there somewhere. Many Worldview 1 methods are derivatives of brainstorming, which is 'quantity leads to quality' thinking in action. Whereas Worldview 1 is an understanding of reality based on the thinking of Newton and Descartes, first order change and the laws of cause and effect, Worldview 2 represents an altogether different view of reality. Under Worldview 2, problem solving, development, innovation and change come about through small nudges that subtly alter the natural flow of events, such that the needs and interests of all relevant parties are satisfied, quickly and with existing resources. Although causal analysis is still useful for solving simple problems, complex issues are addressed through context manipulation, pattern analysis and constraint removal.

Combining text mining technologies with a unique, collaborative methodology that simultaneously

examines user desirability, technical feasibility, and business viability, our innovation engine enables corporate partners to anticipate disruption and provides them with visibility into ideas and technologies that can affect their markets (Figure 1).

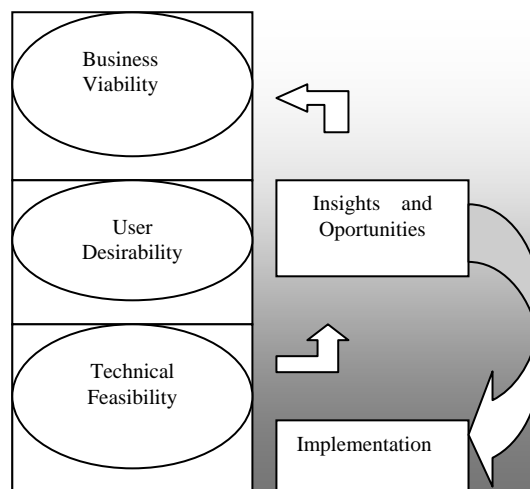


Fig. 1 – Innovation Engine

### 4. METHODOLOGY

The methodology consists of 5 phases as represented in Figure 2.

#### 4.1. Subject Definition:

Like all good science the success of online collaborative research will be about asking the right questions. Communication and the alignment of goals between researchers are vital ingredients to making a multidisciplinary project fly. But clear detailing of project goals, how each participant will contribute and what he or she will gain also are essential.

#### 4.2. Literature Review:

In this phase, the senior academics select published papers containing the most updated scientific and technological developments relative to the subject under study. This phase is accomplished by Web-based paper abstract posting and voting. All the submitted papers are carefully evaluated by each senior academics under the perspective of opportunity and relevance of the contribution, fostering understanding and development of the subject. The evaluation is registered in the form of votes — "relevant" or "not relevant". If the scientist is not familiar to the paper subject he or she may vote for "abstention". Trying to minimize the effect "Rich Get Richer", partial results are not disclosed during the period of voting.

#### 4.3. Insights and Opportunities:

In this phase all submitted papers are processed and clustered via text mining and the results presented to the scientific advisory board for analysis. The board then selects a set of representative papers to serve as base for speculations, refinements and discussions in the next phase.

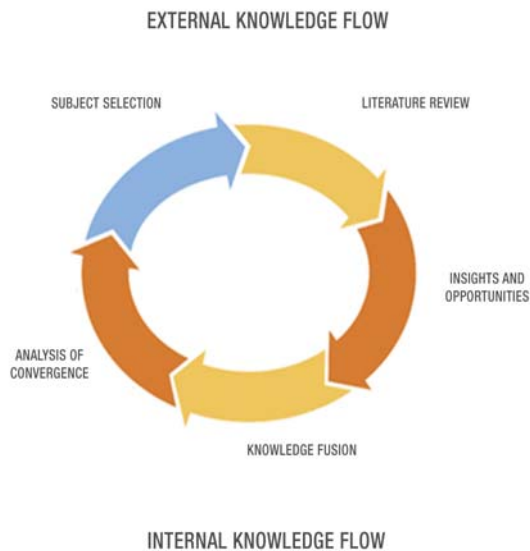


Fig. 2 – Methodology

#### 4.4. Knowledge Fusion:

In this phase the senior academics offer their opinions, ideas and critics, being discussed innovative ideas and solutions posted as comments.

#### 4.5. Analysis of Convergence:

Based on comprehensive text mining analysis, the scientific advisory board may decide to launch a new subject or to ask the senior academics to refine the literature, and/or to continue to discuss the current subject.

Implementation completes the cycle of ideation to drive the concept to its final form.

The possibilities have been evaluated, the prototypes validated and refined. However, innovation isn't something that innovators do — innovation is what customers adopt. It is the market that determines whether something is innovative or not, and whether something is adopted or not. Just because an innovation is adopted doesn't mean that it's a successful business innovation. The challenge for business is not just to get customers and clients to adopt the innovation; it's to get them to pay a premium for the innovation. If the proposed solution is approved by Petrobras, a project team performs detailed design and engineering, chooses manufacturing partners if necessary, and works with the client to perform a timely and successful launch.

### 5. TEXT MINING IMPLEMENTATION

An alert reader will make connections between seemingly unrelated facts to generate new ideas or hypotheses. However, the growth of published articles means that even the most avid reader cannot hope to keep up with all the reading in a field, let alone adjacent fields. New knowledge is at risk of languishing undiscovered in the literature.



Fig. 3 – Web-based Literature Review

Text mining offers a solution to this problem by supplementing the human reader with automatic systems undeterred by the text explosion. It involves analyzing a large collection of documents to discover previously unknown information. The information might be relationships or patterns that are buried in the document collection and which would otherwise be extremely difficult, if not impossible, to discover. The convergence analysis is based on text mining developments, and used to process the abstracts of the

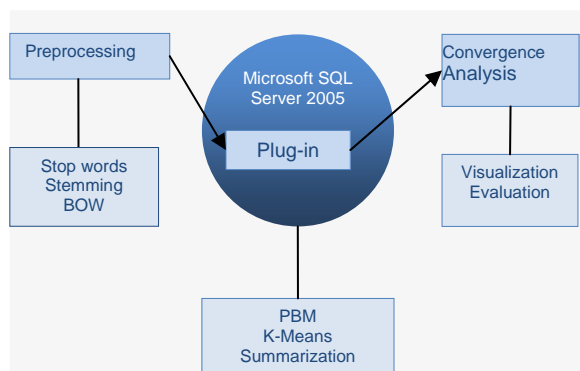
selected papers to start the discussion and to analyze the comments in the knowledge fusion activity (Camarinha-Matos 2005, Zanasi 2005, Liu 2006, MacLennan and Tang 2005, Chickering and Iver 2004). It is completely integrated to the environment, and the diagram of this process is indicated in Figure 4.

To achieve the tight coupling of Data Mining (DM) techniques in Database Management Systems (DBMS) technology, a number of approaches have been developed in the last years. These approaches include solutions provided by both company and academic research groups.

Toward this objective, the Microsoft (MS) Object Linking and Embedding Database for DM (OLE DB DM) technology provide an industry standard for developing DM algorithms. This technology was first included in the MSSQL 2000 release. The MSSQL Analysis Services (SSAS) has included a DM provider supporting two DM algorithms (one for classification by decision trees and another for clustering) and the DM aggregator feature made possible for developers and researchers to implement new DM algorithms. The MSSQL 2005 version has included more seven algorithms as well as a new way to aggregate new algorithms, using a plug-in approach instead of DM providers. This plug-in approach was adopted to build a new text mining environment, and can support 2 languages: English for the papers and Portuguese for the scientist comments.

Preprocessing includes stop words and stemming processing and the BOW (bag-of-words) toolkit, a library of C functions useful for writing statistical text analysis, language modeling and information retrieval programs.

Clustering is accomplished by the well known K-Means (Zanasi 2005, Liu 2006) algorithm and the search of better clusterization process employs PBM index, a product of three factors, of which the maximization ensures that the partition has a small number of compact clusters with large separation between at least two of them.



**Fig. 4 – Text Mining Implementation**

PBM index is calculated by:

$$PBM(K) = \left( \frac{1}{K} \cdot \frac{E_1}{E_k} \cdot D_k \right)^2 \quad \text{where,} \quad (1)$$

$$E_k = \sum_{k=1}^k E_k$$

is the total intra-cluster distance for K clusters.

The third factor is the biggest inter-centroid distance. Automatic text summarization is the technique where a text is entered into the computer and a non redundant extract from the original text is returned.

The technique has its roots in the 60's and has been developed during 30 years, but today with the Internet and the Web the technique has become more important. Microsoft Word has since 1997 a summarizer for documents.

In order to generate a summary, we have to identify the most important pieces of information from the document, omitting irrelevant information and minimizing details, and assemble them into a compact coherent report. This however, is easier said than done as it involves some of the main problems of natural language processing. To produce a domain-independent system would require work in natural language understanding, semantic representation, discourse models, world knowledge, and natural language generation. Successes in domain-independent systems are few and limited to identifying key passages and sentences of the document. More successful systems have been produced for limited domain applications.

The used technique for automatic text summarization has been developed for many years: text extraction, which means to extract pieces of an original text on a statistical basis or with heuristic methods and put them together to a new shorter text with as much information as possible preserved. One important task in text extraction is topic identification. There are many methods to perform topic identification. One is word counting at concept level that is more advanced than just simple word counting; another is identification of cue phrases to find the topic (Edmundson 1969).

## 6. RESULTS

The first Latin America's e-science initiative was tested against reality in their very first iteration. The initial subject was Renewable Energy with Social Responsibility. Using the newly developed online collaboration tool were posted 164 scientific papers, 1,688 evaluations, and 314 comments by 50 senior scientists selected from among the nation's top research institutes. Interestingly, despite the potential for solar and wind electricity generation in Brazil, the multidisciplinary team of academics was not motivated by those alternatives and they started discussing biodiesel production considering microalgae and pollution effects. The innovation engine grouped discussions on 3 clusters: biodiesel, hydrogen and ethanol. At the end of the process 5 research projects were presented to be considered for industrial

development. This was considered as an excellent result, because completely different solutions were proposed.

At this time, as the learning process achieved a good performance, more specific themes are being examined with short time duration cycles.

## 7. CONCLUSION

The research enterprise increasingly involves multidisciplinary collaborations, sometimes over geographic distance. Technological advances have made these collaborations possible, and the history of past innovations suggests these collaborations are highly desirable. The accumulated body of empiric knowledge and the size of the involved research community provide the important conditions for the foundation of a new scientific organization. This paper introduced a view of an application in this area. As important as collaborative tools, techniques and technologies are, you have to have an environment where there are incentives to collaborate. You have to have an internal economy where there are appropriate rewards and incentives for collaborating, and appropriate disincentives for not collaborating. Interestingly enough, we don't always need to collaborate. You have to have a circumstance where you cannot succeed, cannot get promoted in the organization, cannot get extra organizational resources, unless you are deemed to be somebody who is a creative and cost-effective collaborator across the enterprise. The fact of the matter is that many people have succeeded in organizations not because they're very good collaborators, but they've been successful as leaders on their own, they're not team players at all. But just because you have a common goal doesn't mean you actively and appropriately collaborate with your teammates. It takes shared space to create shared understandings and they realized they couldn't create this value on their own. That's what collaboration has to be about, and it must be recognized explicitly and you must be rewarded explicitly.

## ACKNOWLEDGMENTS

This research has been supported by Petrobras S.A. the Brazilian Energy Company.

## REFERENCES

- ACM, A new interview with M. Schrage , Available from:  
[http://www.acm.org/ubiquity/interviews/v7i08\\_schrage.html](http://www.acm.org/ubiquity/interviews/v7i08_schrage.html) [accessed 5 July 2008]
- ACM, 2006, Service Systems, Service Scientists, SSME, and Innovation, Communications of the ACM, Vol. 49, No.7, 81--84, July .
- Camarinha-Matos, L.M. and Afsarmanesh, H., 2005, Collaborative networks: a new scientific discipline, Journal of Intelligent Manufacturing, 16, 439--452.

- Chickering M. and Iver, R., 2004, A tutorial for constructing a plug-in algorithm, Microsoft Technical Articles.
- Edmundson H. P., 1969, New Methods in Automatic Extraction, Journal of the ACM 16(2) pp. 264--285.
- Gorman, M., 2007, Web 2.0: The Sleep of Reason Part II, Available from:  
<http://blogs.britannica.com/blog/main/2007/06/web-20-the-sleep-of-reason-part-ii/> [accessed 5 July 2008]
- IAC, 2007, Lighting the way: toward a sustainable energy future, InterAcademy Council, October, in:  
<http://www.iafastro.com/index.php?id=498> [accessed 5 July 2008]
- Iver R. and Crivat, B., 2004, SQL Server data mining plug-in algorithms, Microsoft Technical Articles.
- Lanier, J., 2006, Digital Maoism - The Hazards of the New Online Collectivism, Available from:  
[http://www.edge.org/3rd\\_culture/lanier06/lanier06\\_index.html](http://www.edge.org/3rd_culture/lanier06/lanier06_index.html) [accessed 5 July 2008]
- Lee, P., 2006, Convergent technologies: Future perfect or imperfect?, Available from:  
[http://www.wacc.org.uk/wacc/publications/media\\_development/2006\\_2/convergent\\_technologies\\_future\\_perfect\\_or\\_imperfect](http://www.wacc.org.uk/wacc/publications/media_development/2006_2/convergent_technologies_future_perfect_or_imperfect) [accessed 5 July 2008]
- Liu, B., 2006, Web Data Mining Exploring Hyperlinks, Contents and Usage Data, Springer, December.
- MacLennan, J. and Tang, Z., 2005, Data Mining with SQL Server, Wiley Publishing Inc.
- Petrobras, 2008, Available from:  
<http://www2.petrobras.com.br/ingles/index.asp> [accessed 5 July 2008]
- Tang, Z. and MacLennan, J., 2005, Data Mining with SQL Server 2005, Wiley Publishing.
- Zanasi, A., 2005, Text Mining and its Applications to Intelligence, CRM and Knowledge Management, WIT Press.