Internet Search mechanisms and distortions of the semantic space: The Scientific challenges facing the "googles"

Alexandre Linhares

Fundação Getulio Vargas

Email: linhares@clubfrome.org.br

Carla Winter Afonso

Fundação Getulio Vargas

Email: cwinter@clubfrome.org.br

Abstract

Ever since the launch of Altavista, internet search engines have become a multi-billion dollar industry, with fierce competition between Google and the three major competitors. One of the challenges involved is to rank search results in a way that places the most meaningful results at the top. In order to do this, the algorithms involved must try to grasp the actual meaning, the semantics, embedded in a search query. In this paper we discuss a problem we call "distortions of semantic space". Distortions of semantic space occur regularly in people's texts, writing styles, labeling of images, etc. We present a number of examples of distortions of semantic space, and analyze the problem. We also comment on new computational architectures that have tried to handle this problem, albeit the state of the art still remains far from the needed.

Key words:
Search mechanisms, distortions of the semantic space, literal search

"If someone has broadband, dial-up, or access to an internet café, whether a kid in Cambodia, the university professor, or me who runs this search engine, all have the same basic access to overall research information that anyone has. It is a total equalizer. This is very different then how I grew up. My best access was some library, and it didn't have all that much stuff, and you either had to hope for a miracle or search for something very simple or something very recent. [...Google gave that kid] universal access."

Sergei Brin, Google Founder

"Previously to Google, when the Secretary of State asked an advisor for a UN Security Council resolution, the advisor would just go get it and bring it to the Secretary. Now, the Secretary of State googles the Security Council resolution and the advisor is better be prepared to interpret and discuss it."

Thomas Friedman, Pulitzer Winner

1. Introduction

Google defines its mission as "to organize the world's information." Since its launch, in 1998, it has reached enormous financial and marketing success, given its superior ranking and indexing technology of data in the Internet. It is now possible to carry searches in 100 different languages with Google, and in 2005, the company reached the mark of *a billion searches per day* (Friedman 2005). To sustain this leading strategic position, however, the company faces enormous scientific obstacles so that, as the types of information available on the web change, new technologies may be able to organize them in an agile form for all to access. The questions with which this paper deals is: *(i) what* *are the main scientific challenges, the basic science obstacles, involved in developing future search engines? (ii) how can these obstacles modify strategies and the positioning of diverse players in this enormous and rapidly expanding market?* (In this paper, the term 'google' is used throughout, but the arguments hold for all search mechanisms.)

1.1. Organizing the world's information

One of the greatest landmarks in the evolution of the Internet was the appearance of search mechanisms such as Google, which quickly succeeded Altavista in market leadership. The gigantic amount of information available on the web was, previously, of difficult access; as sites such as Yahoo! or Internet Yellow Pages (today only of historical value) tried to organize such data using a directory structure, cataloguing each page and site according to the interpretation of their employees. Two problems emerge with this approach:

(i) *The interpretation of the employee who initially catalogued the page* could be different from the interpretation of the user; suppose an employee categorized eBay, the giant auction website, in a /shopping/auctions directory structure. Imagine now that a specific user were searching for a "place to find people who collect stamps". eBay obviously is such a place; however, classification through a directory structure can not lead all its potential user base to it.

(ii) *The scalability of the model*, as the number of pages available grew from a few hundreds, to thousands, then millions, to today's billions. It is not economically viable to pay large amounts of people to catalogue billions of pages, and, even if it were, that would be a *Sisyphus* task, as these pages are in constant content change.

As we will see below, these factors enabled Google to conquer a significant part of the added value in organizing the internet's information.

1.2. Strategic Sustainability: the best results in the top.

Two questions are crucial to understand the success of Google and the sustainability of its strategy. (i) Why is Google the leader of the search market? (ii) What supports Google in that leadership position?

Why does Google lead the market of searches? The first-mover advantage assumption is, in this case, simply wrong, as Google had at least 7 previous mechanisms in the brief history of the WWW:

(i) WWW Wanderer
(ii) WWW Worm
(iii) Webcrawler
(iv) Lycos
(v) Infoseek
(vi) Excite
(vii) Altavista

These two initial engines considered only page headers, and not the pages' main content. Altavista, launched by the research department of Digital Corporation as demonstration of the power of its 64-bits "alpha" processor, was the first engine to consider the entire content of all pages in the Internet – which guaranteed the leadership of Altavista until the launch of Google. Unhappily for Digital Corp., Google possessed basic characteristics that would enable it to quickly surpass Altavista. *These characteristics are the target of our* *work, and will be dealt with in section 2.* Today the search mechanisms that divide market share are:

(i) Google
(ii) Ask (formerly Ask Jeeves)
(iii) MSN Search (Microsoft)
(iv) Yahoo! Search

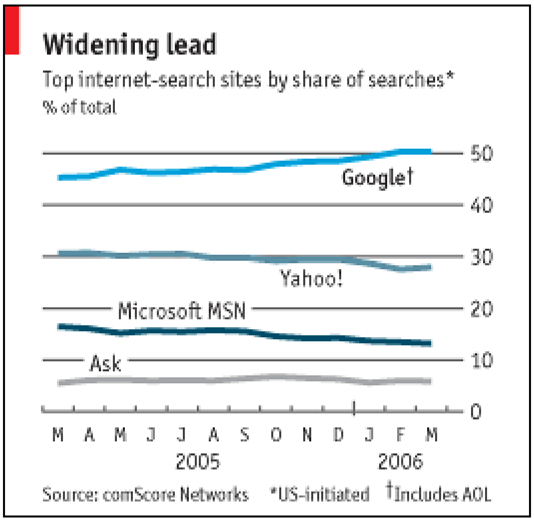

Figure 1. *With over 50% of the search market, Google remains in an absolute leadership position (The Economist 2006).*

The second question involved is: what supports the company in this position of market leadership?

Because it is based on technological standards, which demand high R&D costs to establish, and later become a formidable barrier to entry, the technological industry historically was dominated by a leader, who defined the market for other companies. During the 60's up to the 80's, IBM dominated the market, and the other actors developed their strategies accordingly. After the launch of the IBM PC, value migrated to IBM's suppliers: processors (Intel) and operation systems (Microsoft) dominated the architecture of the PC era, and as 'clones' became widespread, IBM could not dominate this new market. Microsoft, specially, became a giant with the capacity to jam competitors: Companies such as Wordperfect, Lotus, Netscape, and DR-DOS were once important, but today few remember these names. The journal *The Economist* raises, then, an important point: is Google the new Microsoft?

The comparison is both a compliment and a reproach. It is a compliment because it implies that Google has now become the company that defines the environment in which other technology firms operate, just as IBM and Microsoft once did. As with Microsoft in its heyday, Google is the technology firm where the smartest geeks aspire to work; it embodies the technological zeitgeist; and it is a highly regarded company that has become a household name. But the comparison is also a reproach, because it highlights growing concern that Google is now too powerful for its own good, or that of the industry, or indeed that of the world at large.

As *The Economist* (2006) points out, despite all the similarities between these technological leaders, Google's strategic position is not as sustainable as that Microsoft once held. While Microsoft possessed architectural standards, first with MS-DOS, then with MS-Windows, and used its clout to establish MS-Office as a de facto standard, the switching costs to rival architectures deemed, per user, (i) a high learning effort, (ii) an effort to transfer and convert files, and (iii) an effort to constantly share with others files that must be compatible with Microsoft's standard. Although now Microsoft is seriously threatened on diverse fronts, such as the movement of open software that includes operation systems (such as Linux), web browsers (such as firefox) and applications (such as OpenOffice), the switch to a new computing architecture still brings per user costs that tend to multiply when the switcher is a corporation with a large installed user base.

In contrast, the Google user remains only one click away of its competitors, such as Ask, Yahoo, or MSN search. Therefore, referring specifically to search systems, a Google user possesses minor costs of transition to a new service provider -- which makes the company more vulnerable to competitors, and obliged to keep highly relevant results; which, technically, means to keep its index better organized than the competition's.

1.3. The value of the service

What is the value of Google services? One of the forms to evaluate the company is to verify its financial market value. Recent fluctuations of Google stocks are presented below in Figure 2. After reaching a maximum above USS470, the value had fallen, in May of 2006, to USS370. This means that the value of the company as a whole would be above USS113.000.000.000,00 (113 Billion dollars). As a comparison, the biggest Brazilian company, Petrobras, was, also in May 30th 2006, evaluated at 99 Billion dollars.

Figure 2. *Google's market capitalization based on its stock value surpasses 100 Billion dollars. The company, who approximately*

possesses revenues of six billion dollars, is evaluated as more valuable than Petrobras, whose revenues surpass 50 billion dollars (the reader should also consider that the Petrobras stock value has grown considerably due to increases in international prices of crude oil).

2. "Intent drives search": from psychology to new mechanisms of search

"Search is a problem 5% solved", says Udi Manber, the CEO of the search mechanism A9 from Amazon.com (Batelle, 2005). In this section we explore the nature of the search problem. Not all searches are for a determined topic of a subject (Batelle 2005). Approximately 15% of the searches look for a good set of links, in contrast to a good document. Approximately 25% of the searches are *navigational*, that is, for a specific website that the user already had in mind. About 36% of the searches are made with sights to a transaction, either commercial, or information on tracking of a package, etc. Approximately 12% of the volume of searches is referring to sex. Given this variety of initial intentions, we can start to visualize why the problem is only 5% solved (Battelle 2005). Engines still have to consider that the common user generally type only one or two terms in a search.

Let us assume, for example, a search for the word "jaguar". Which type of pages must appear as the first ones? Consult the word Jaguar on Ask.com, and it will guide the user for a definition of the *original intention*: "animal jaguar"?; "Car"?; etc. Our mind is extremely fast in processing ambiguous information and reinterpreting them within the context (Hofstadter 1995; Linhares 2001). As example, the ambiguous phrase "prostitutes appeal to pope" is mentally reorganized after an initial interpretation, but engines lack this reinterpretation capability. A search for "biography abraham Lincoln" does not mean that a desire for all biography pages mentioning "Abraham Lincoln".

Symbols and semantics: What is it really desired from one or two requested words? Given one or two symbols, which is the meaning that you looked for? If one wants to understand what Manber had in mind when he said that the problem is 5% solved, we can observe a quote from Batelle (2005), mentioning the problem of understanding original intention:

But how might we get there? For search to cross into intelligence, it must understand a request--the way you, as a reader, understand this sequence. [...] My problem is understanding something. That can only happen if search engines understand what a person is really looking for, and then guide them towards understanding that thing, much as experts do when mentoring a student.

This problem seems simple, yet, it is daunting. Consider, for instance, the question "What is similarity?", as it applies to text documents such as those indexed by search engines. If Google found two documents with thousands of words in exact sequence but a mere comma of difference between them, should the engine classify such pages as similar? It seems obvious, for there is no reason the algorithm might dismiss a mere comma to make any significant change in what concerns the content of the documents. Then again, consider it from a human's eyes.

This is an intriguing example, from a story reported on the New York Times (Ian Austen, The Comma That Costs 1 Million Dollars [Canadian], October 25, 2006):

The Comma That Costs 1 Million Dollars (Canadian)

OTTAWA, Oct. 24 — If there is a moral to the story about a contract dispute between Canadian companies, this is it: Pay attention in grammar class.

The dispute between Rogers Communications of Toronto, Canada's largest cable television provider, and a telephone company in Atlantic Canada, Bell Aliant, is over the phone company's attempt to cancel a contract governing Rogers' use of telephone poles. But the argument turns on a single comma in the 14-page contract. The answer is worth 1 million Canadian dollars ($888,000).

Citing the "rules of punctuation," Canada's telecommunications regulator recently ruled that the comma allowed Bell Aliant to end its five-year agreement with Rogers at any time with notice.

Rogers argues that pole contracts run for five years and automatically renew for another five years, unless a telephone company cancels the agreement before the start of the final 12 months.

The dispute is over this sentence: "This agreement shall be effective from the date it is made and shall continue in force for a period of five (5) years from the date it is made, and thereafter for successive five (5) year terms, unless and until terminated by one year prior notice in writing by either party."

Consider that last comma. How long should the contract last? Without the comma, it's pretty clear, right? It must last at least a full 5 years. It is beyond the point whether the lawyers actually intended this, but the comma, however, distorts meaning in a profound way. This distortion of meaning brought by the slightest of cues is a significant cognitive phenomena, for it happens, many times, subconsciously in a human’s information-processing, with no need for any conscious thought.

Let us now get back to Google's way of looking at things. There are two 14-page documents, one has a single comma that the other lacks. Should Google classify them as "similar"? It seems clearly obvious that it must be the case: to Google's eyes, these are 99,9999% similar. After all, under what circumstances should the algorithms in a search engine perceive the semantic dangers that lie within a single comma, given thousands and thousands and thousands of exactly-matching-words-and-paragraphs documents?

3. Focusing the problem

In this section we discuss the nature of some problems regarding search mechanisms. We initially consider the problem of literal search, and later, the problem of search for multimedia content of dauntingly difficult indexation.

3.1. Literal search

In 1957, a thought by J.R. Firth launched an idea well used in the study of linguistics, which later would influence the mechanisms of literal search: "*You shall know a word by the company it keeps*".

Behind this phrase is the idea of *correlations between words* that help understand the meaning inherent to each word. Words with similar meanings would tend to appear in a great number of texts, and, therefore, its meaning could be extracted from the analysis of the relations between words. In fact, this was the idea used in search mechanisms. This seems to be a simple mechanism for extracting intent, yet, we claim that the mechanisms of literal search face four basic problems:

(i) Deformations of the semantic space - similar words are considered next in the semantic space. Through the process of analogies we perceive an object as pertaining to another class of objects. An mp3 player, of Apple, iPod, can be seen as "walkman", but also it can be seen as "a printer", or "ferrari", or a "Trojan horse" (Afonso and Linhares, 2007). Another example given by French (2002): the word "hammer" is next in meaning to saw, nails and other construction materials, but one is capable of attributing different meanings to the same objects. The hammer can as a paperweight, losing its initial function (and starting to become related with different objects) in semantic space. Linhares and Brum (2007) have shown that this effect arises in chess players strategic thinking.

(ii) The mechanisms of literal search do not detect the occurrence of abstract structures - through the process of analogies we compare different things: an iPod "is a Ferrari of mp3 players"; "Google is the new Microsoft", etc.

(iii) The mechanisms of literal search do not know the words in the same way we do - we know the words through experience and contact with the world, which makes them assume multiple meanings and connotations to us; but not to search engines.

(IV) They consider that words are atomic entities- for human beings words are not atomic; therefore syllables can assume distinct functions, which complete the meaning of the word.

It is due to these four basic problems that the thesis of correlation between words helping to understand the meaning inherent to each word may be discarded. We can see that in some examples: when we ask the system “a good name for Father”, the word "John" appears, obviously. But the word "Mary" also appears. More interesting: when we ask “a good name for the prime minister of Israel”, the words "Sharon", "Isaac", "Rabin", appears in the top. But also appears "Arafat”. Why? Because the searches are made based on correlations, and Arafat obviously is correlated to "prime minister of Israel" in millions of texts in the web.

The systems of literal search are blind for certain connections that we make easily. Hofstadter (1995) discovered that our mind is only capable of understanding things because it perceives, impulsively, subconsciously, abstract roles for words and things; therefore we use so many analogies. When we ask the system to classify "how much you perceive lawyers as":

(i) telephones
(ii) sharks
(iii) blood suckers
(iv) vampires
(v) rocks

The system says that lawyers are more "telephones" than "vampires" or "bloodsuckers", when most people respond otherwise. Why does the system make such erroneous mistakes? Because it is blind to the abstract roles that we see lawyers portraying in our society. The system is incapable of making analogies that we make immediately. What we, human beings, see, when we understand what we see, are abstract roles that allow us to make analogies (Linhares 2005; Hofstadter 1995). Let us see some examples:

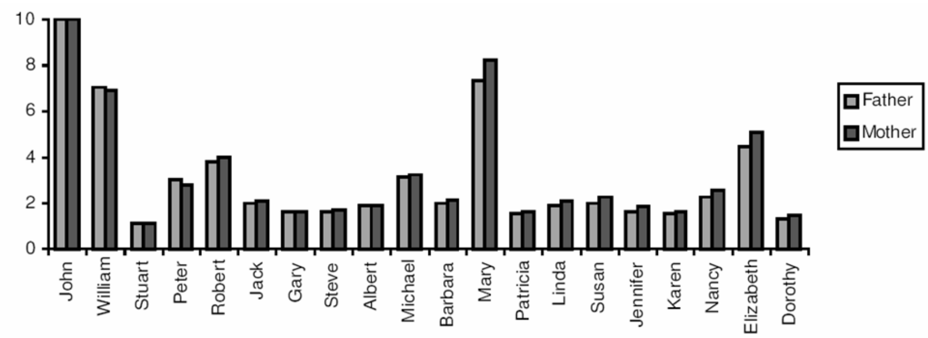

Figure 3. *What is a good name for a "Father" ? What is a good name for a "mother"? As these words tend to appear in similar contexts (example: "the mother of Jack"), the results are very similar for both sexes.*

(In: French, R. M. and Labiouse, C. (2001). Why co-occurrence information alone is not sufficient to answer sub cognitive questions. Journal of Theoretical and Experimental Artificial Intelligence, 13(4), 419-429)

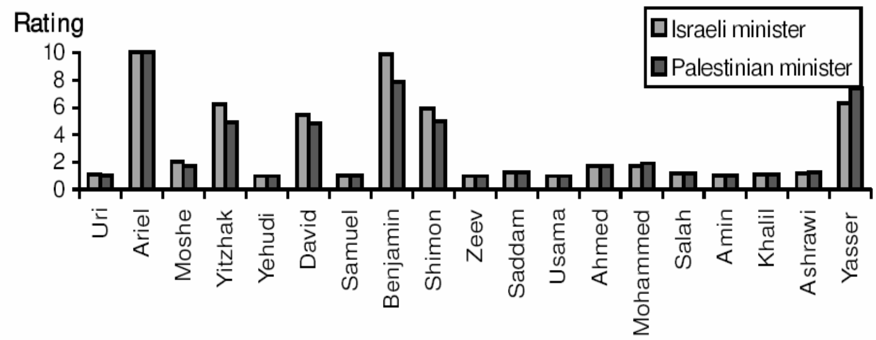

Figure 4. *Which is a good name for prime minister of Israel? Which is a good name for prime minister of Palestine? As in the example above, the proper names are correlated with both sides, so that Saddam Hussein seems a good name for prime minister of Israel ("prime minister of Israel threatened Saddam Hussein...").* (In: French, R. M. & Labiouse, C. (2002) . Four Problems with Extracting Human Semantics from

Large Text Corpora. Proceedings of the 24th Conference of the Cognitive Science Society.)

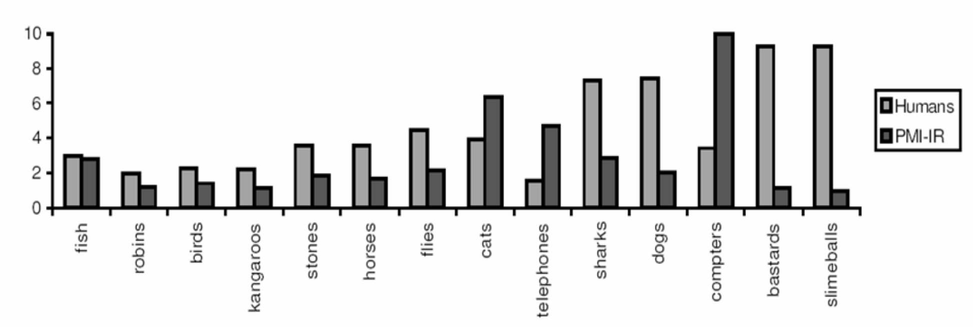

Figure 5. *When we asked the system to "rate lawyers as: horses, fish, telephones, stones, sharks, cats, flies, birds, slime balls, kangaroos, robins, dogs, and bastards", the results are the opposite of what humans think.*

(In: French, R. M. (1997). When coffee cups are like old elephants or Why representation modules don't make sense, Proceedings of the International Conference New Trends in Cognitive Science, A. Riegler & M. Peschl (eds.), Austrian Society for Cognitive Science, p. 158-163.)

The skeptical reader could argue: "does this type of anomaly occurs in practical situations? Could a system such as Google really offer this type of results? "Let us see one example of the following search: "Israeli prime minister name" (carried through in May 30th of 2006). Between ' top ten hits ', we can find:

BBC NEWS | Middle East | Hamas 'names its prime minister' **]** Israel says it will not deal with a Hamas government unless it renounces violence ... We have decided to nominate brother Ismail Haniya as prime minister ...**news.bbc.co.uk/2/hi/middle_east/4721456.stm - 41k**

It is indeed the case that search engines are 'fooled', and point out exactly the enemies of those intended in the search query! The goal of this paper is, therefore, to detail and explain what are the imperfections of the system related to the original intention of the user. This task is very important when you talk about indexing words, but it still gains more importance in the ever changing context of the internet.

3.2. The new contents of the Internet and the failure to index them

Innovative new media appear regularly in the web, such as podcasts, video casts, images, etc. The content of pages in the web, which in the past consisted of simple texts and images become increasingly more complex. While words can be easily catalogued, the content of the net is quickly migrating to a type that will be composed of data of difficult indexation. For example, if one wants to find a podcast that "comments the conflict in Iraq from the perspective of the allied soldiers who are against the whole ordeal", will be extremely difficult to find such without intense effort. Obviously there are innumerable podcasts commenting the conflict, and there is high probability that at least one will comment from the perspective of unsatisfied allied soldiers. Unhappily, this type of information is of difficult indexation (even with the use of tags), which makes it a daunting challenge to organize the information in the net as it becomes enormous -- and, as we saw, incredibly valuable.

As an example, a simple search for images can generate atypical results, not to say comic. Below are some examples of search for images that demonstrate its intrinsic difficulty of indexation. Let us consider three searches in the following format: [adjective] [substantive].

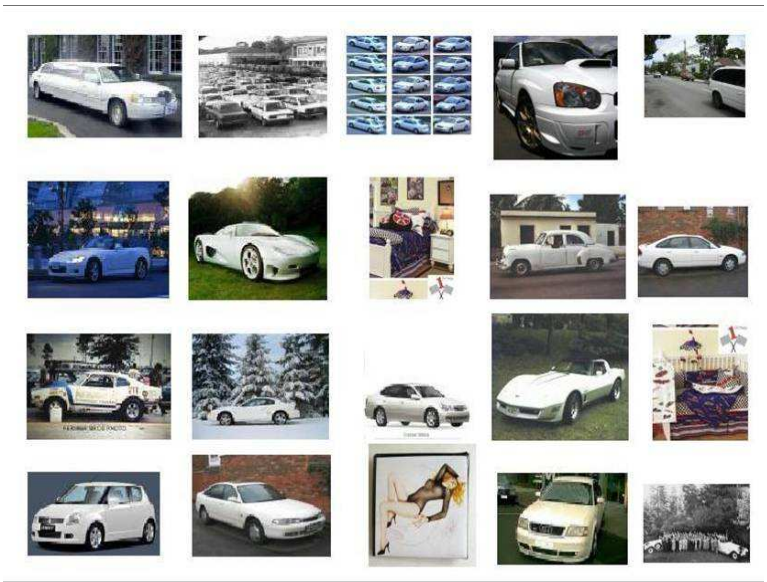

Figure 6. *Search for "White car" carried through in google image search. Given the images to a classroom with 40 students, they find in seconds which was the word used in the search. Note that some images are absolutely irrelevant to the intended query.*

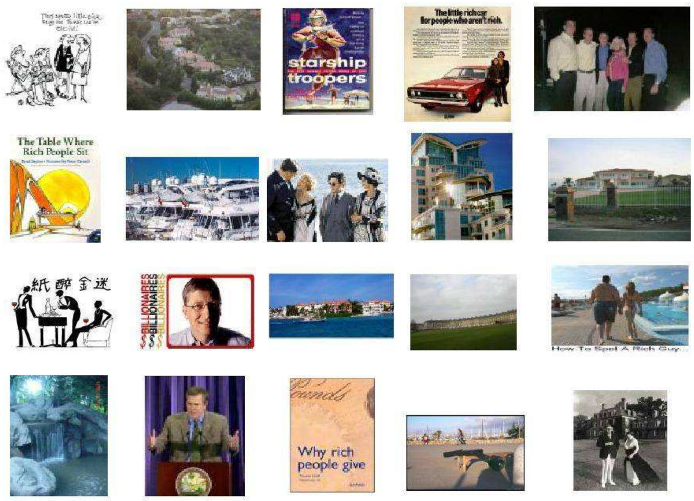

Figure 7. *Search for "Rich people" carried through in google image search, where the results start to become strange. Given the images to a classroom with 40 students, they need some minutes, and some hypotheses, to find the word used in the search.*

Figure 8. *Examples of images badly classified by google image search. Search for "Disappointed firefighter" where the results are unrecognizable, given the original intention. Given the images to a classroom with 40 students, they are incapable to find the word used in the search, after dozens of different hypotheses.*

4. Objectives of the Study

The purpose of the paper is to identify main obstacles of basic science and how these obstacles can modify the strategy, and the positioning, of diverse players in this rapidly expanding market. As intermediate objectives we identify:

1- Identify the problems of literal search;
2 - To search on the new contents of Internet and the failure to index them, developing following sub-items:

(i) How to make a search for a video where "somebody is about to be surprised"? It is impossible, today, given the abstraction of the concepts "surprise", "about to be", etc. Before having such information catalogued, we need to make machines understand videos.

(ii) How to understand images? How to perceive a man in an image? How to perceive that he is a man, and not military woman, with man's clothes and short hair? How to perceive that he is a man and not a picture (or a photo)? How to make a search for "an image where somebody is being ridicularized , but still seems happy" ?
(iii) These lead us to Bongard problems, how to understand subtle differences that exist even in simple geometric figures. We need to solve the problem in general. But today, it is not possible to classify automatically the Bongard problems (Linhares 2000). The system is not capable of distinguishing "great figures from small figures" and at the same time "slinder figures versus total scribbled figures", and at the same time "three objects versus four objects" (see Linhares 2000).

5. Conclusion

This paper considers the analysis of the main involved scientific obstacles in the development of search mechanisms, in a rapidly expanding market. We criticize the subordinated vision of the correlations involved in the search mechanisms. Finally, we comment on a new movement known as the Semantic Web.

The Semantic Web is a mesh of information linked up in such a way as to be easily processed by machines, on a global scale. The Semantic Web was invented Tim Berners-Lee, the father of the WWW, URIs, HTTP, and HTML. The Semantic Web does not have as objective to train machines to behave as people, but to develop technologies and languages that make information legible for machines. The goal is to develop a technological model that allows global sharing of knowledge through the use of machines. Tim Berners- Lee expressed his vision as follows:

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize.

The Semantic Web has as main purpose to attribute one meaning (direction) to the contents published in the Internet in a way that is perceivable for the human being and for the computer. The integration of languages or technologies, architectures of metadata, computational ontologies, agents, among others, will favor the appearance of services that guarantee the cooperation.

James Handler (2001), says that the Semantic Web, in the beginning, will be formed by "knowledge islands", or either, specific niches of knowledge for some application but that, through interoperability between ontologies they will be able to interact.

We assume that the computer understands this phrase:

Anakin Skywalker is Luke Skywalker's father

For human beings it is simple what the sentence means - Anakin and Luke are people and a relation exists between them. You know that father is a type of kinship and also knows that this sentence means that Luke is son of Anakin. But the computer does not understand this sentence without an aid. To allow the computer to understand what the sentence means, you need to include information that describe who Anakin and Luke are and which is their relation. For this function two tools had been developed: eXtensible Markup Language (XML) and Resource Description Framework (RDF).

XML is a language as HTML that governs the appearance of the information in the Internet.

The XML adds information to the HTML, including information that describes the data. This information is invisible for the reader of the text but accessible for the computers. RDF works separating the phrase in the following parts:

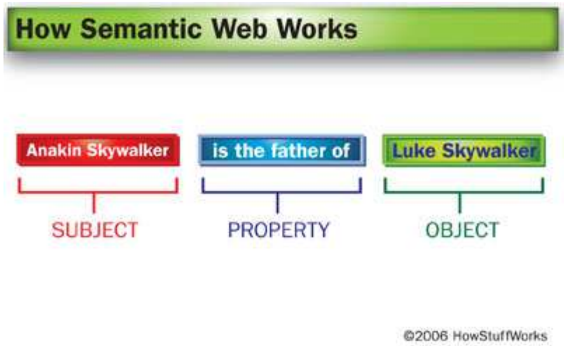

Figure 9. Model of the *Semantic Web.*

The computer perceives that two objects in the sentence exist and a relation between them. But the computer still does not recognize what are the two objectives and how they become related. Another tool, uniform resource identifiers (URIs) allows the computer to access the information on objects and the relation between them.

The Semantic Web possesses a set of ontologies that allows the understanding of words. We display the ontologies used by the Semantic Web:

- RDF Vocabulary Description Language schema (RDFS) - RDFS adds classes, subclasses and properties to the words. Example: Dagoba, name is a class, a planet, subclass.
- Simple Knowledge System Organization (SKOS) - SKOS classifies the words in including or restrictive way. Example: Sith Lord is classified in restricted way as Darth Sidious and in an including way as villains.
- Web Ontology Language (OWL) - OWL describes relation between the classes and uses the logic to make deductions. It can also construct new classes in the existing information.

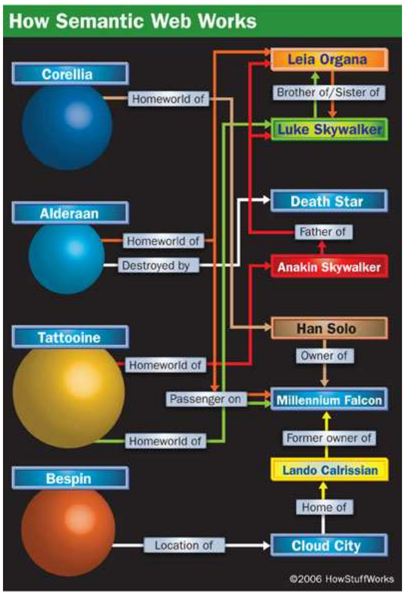

Figure 10. *Architecture of relations between information of the Semantic Web.*

Companies such as IBM, Microsoft, among others, are intensely investing in the "Semantic Web". The intent reader already must have perceived that the proposal of the Semantic Web also suffers from the same imperfections previously discussed. The involved mechanisms do not consider the fast distortion of the semantic space, and, therefore, they are incapable of understanding simple phrases such as "that doctor is a butcher", in which an analogy distorts the meaning involved. With the use of XML and RDF, the Semantic Web would tend to interpret that "that doctor" possesses another occupation, another work, and could not understand the pejorative way people may talk and refer to it. For these reasons we propose that our discussion is of ample applicability to the study of the Semantic Web (see Linhares and Brum 2007, Linhares 2000, Linhares 2005, Hofstadter 1995).

Referências:

The Economist (2006) *Is Google the new Microsoft?*, Leaders column, Information technology, May 11th

Afonso, C. W. Linhares, A. (2007). Analogias no processo decisório. *Working paper.*

Battelle, J., (2005) **The Search: how Google and its rivals rewrote the rules of business and transformed our culture.** New York: Penguin Books.

Calishain, T., e Dornfest, R. (2005) **Google Hacks.** Sebastopol: O'Reilly.

H. R. Ekbia (1993) **AI Dreams and Discourse: Science and Engineering in Tension**. Ph.D. Thesis, IU Bloomington.

French, R. M. and Labiouse, C. (2001) . Why co-occurrence information alone is not sufficient to answer sub cognitive questions. **Journal of Theoretical and Experimental Artificial Intelligence***, 13(4)*, 419-429.

French, R. M. & Labiouse, C. (2002) . *Four Problems with Extracting Human Semantics from Large Text Corpora.* Proceedings of the 24th Conference of the Cognitive Science Society.

French, R. M. (1997). *When coffee cups are like old elephants or Why representation modules don't make sense*, Proceedings of the International Conference New Trends in Cognitive Science, A. Riegler & M. Peschl (eds.), Austrian Society for Cognitive Science, p. 158-163.

Friedman, T. (2005) **The world is flat**. New York: Farrar, Straus and Giroux.

Handler, J. Handler, J., Berners – Lee, T. & Lassia, O.(2001) The semantic web. *Scientific American*, 284(5):34–44, May.

Hofstadter, D. (1995) **Fluid Concepts and Creative Analogies**. New York: Basic Books.

Linhares, A. (2000) *A Glimpse at the Metaphysics of Bongard Problems*, **Artificial Intelligence** 121, pp. 251–270.

Linhares, A. (2005) *An active symbols theory of chess intuition*. **Minds and Machines** 15, pp. 131--181.

Linhares, A. and Brum, P. (2007) Understanding our understanding of strategic scenarios: what role do chunks play? Accepted for publication, **Cognitive Science**.

Smith, B. (1996) *On the origin of objects.* MIT Press.

Standage, T. (Ed.) (2005) **The future of technology.** London: The Economist in association with Profile Books, Ltd.

Vise, D.A., e Malseed, M. (2005) **The Google Story: Inside the hottest business, media and technology success of our time.** New York: Bantam Dell.