

AS MÚLTIPLAS SINTAXES DOS MECANISMOS DE BUSCA NO CIBERESPAÇO

LAS MÚLTIPLES SINTAXIS DE LOS MECANISMOS DE BÚSQUEDA EN EL CIBERESPACIO

Silvana Drumond Monteiro - drumond@sercomtel.com.br

Professora Adjunta do Departamento de Ciência da Informação, Universidade Estadual de Londrina.

Doutora em Comunicação e Semiótica.

Pesquisadora PQ 2 do CNPq.

Líder do Grupo de Pesquisa Informação e Conhecimento no Ciberespaço.

Resumo

No período pós-moderno as máquinas estão organizando o conhecimento por meio dos mecanismos de busca. Para discutir e conhecer melhor essa realidade, o artigo propõe uma tipologia para os mecanismos de busca, visando a ilustrar as múltiplas sintaxes de organização e busca do conhecimento e informação no ciberespaço. Dessa forma, apresenta as categorias de classificação: anatomia, que são os processos básicos de funcionamento desses buscadores, ou seja, *crawling*, *indexing* e *searching*; a forma geral de indexação ou organização, que são os diretórios, programas, metabuscadores e híbridos; a ordenação e apresentação dos resultados, que refletem as múltiplas sintaxes; e o paradigma semiótico, que ilustra a multiplicidade semiótica dos mecanismos de busca. A partir dessa tipologia, os buscadores são exemplificados e discutidos, abarcando esse cenário de múltiplas sintaxes de indexação e busca da informação e conhecimento no ciberespaço e ao mesmo tempo construindo uma ontologia desse objeto. O ciberespaço sempre terá e trará o problema da taxonomia do conhecimento e da multiplicidade dos signos, seja em sua representação ou organização.

Palavras-Chave

Tecnologias da Informação e Comunicação. Mecanismos de Busca. Ciberespaço. Sintaxes. Semiótica.

1 INTRODUÇÃO

Pensar na organização do conhecimento na atualidade significa compreender como as máquinas estão operando a indexação no ciberespaço.

Esse fenômeno, decantado na Filosofia como o pós-moderno, abrange as mudanças ocorridas nas ciências, nas artes e nas sociedades avançadas desde 1950. “O pós-modernismo é coisa típica das sociedades pós-industriais baseadas na informação.” (SANTOS 2006, p. 11).

O termo é de ampla definição e foi cunhado pelo historiador britânico, também filósofo, Arnold Joseph Toynbee, na década em 1940, quando escrevia seus 12 volumes intitulados “Um estudo da História” (COSTA, 2007).

Para Lyon, há distinção entre o “pós-modernismo” e “pós-modernidade”, sendo que a ênfase do primeiro recai sobre as manifestações culturais e intelectuais e do segundo sobre a sociedade em geral (LYON, 1998).

Dessa forma, algumas características são tributárias a esse movimento intelectual, entre eles, a crise das hierarquias do conhecimento, a emergência do local ou singular em lugar do universal, a substituição do logocentrismo pelo iconocentrismo e a proeminência das Tecnologias da Informação e Comunicação (TICs).

Na Filosofia, o pós-moderno encontra seus precursores na década de 60 e alguns nomes são representantes dessa vertente de pensamento, como Jacques Derrida, Jean Baudrillard e Gilles Deleuze, embora nem todos tenham a mesma visão sobre os signos e as TICs. A visão pessimista, de Baudrillard (1981) é que a pós-modernidade, com seus conceitos de desreferencialização do real e a hiperrealidade, recria um mundo substituído pelos signos, uma verdadeira Semiurgia, a substituição do livro pela TV, a migração da palavra para a imagem, do discurso para a representação.

Para Santos (2006) vários princípios do pós-moderno começam com o prefixo “des”, a saber:

- a) des-referencialização do real;
- b) des-materialização da Economia;
- c) des-estetização da Arte;

- d) des-construção da Filosofia;
- e) des-politização da Sociedade;
- f) des-substancialização do Sujeito.

Não obstante, nem todos os filósofos têm uma visão “esvaziadora” do pós-moderno, embora tenham análise crítica sobre algum regime de signos.

Na década de 90, o ciberespaço é o espaço virtual que alguns conceitos-chave do pós-moderno tornaram-se visíveis e funcionais como, a “desterritorialização” e a virtualização dos signos (DELEUZE; GUATTARI, 1995). São conceitos caros à Linguística e ao estruturalismo, é verdade, mas abre uma gama de nuances interpretativas do novo fenômeno semiótico ocorrido no ciberespaço.

Se a Internet é a rede mundial de computadores, concebendo a infraestrutura técnica da rede, o ciberespaço pode ser considerado como a rede mundial de signos e pessoas. Para prezar por um rigor conceitual, o ciberespaço pode ser definido como:

Uma grande máquina abstrata, porque semiótica, mas também social, onde se realizam não somente trocas simbólicas, mas transações econômicas, comerciais, novas práticas comunicacionais, relações sociais, afetivas e sobretudo, novos agenciamentos cognitivos. [...] Um espaço semântico/ semiótico, onde o signo se dá em várias semióticas, desterritorializado, nômade, em escrita espacializada e com a memória em constante modificação. (MONTEIRO, 2007, p.1; 12).

Vale observar que nossa filiação a Gilles Deleuze e a Felix Guattari se deve por explicar os princípios filosóficos do hipertexto, ciberespaço e, por conseguinte, sua organização. Mas, em especial, porque seus conceitos maquínicos atendem a necessidade teórica do estudo das TICs, presentes na área da Ciência da Informação. Assim, nossa abordagem epistemológica é aquela que põe em conexão a cognição e as TICs para estudo dos mecanismos de busca no ciberespaço. Em outras palavras, os objetos que estudamos são tanto semióticos quanto materiais.

Sobre a evolução da organização do conhecimento, Weinberger (2007) esclarece que a “primeira ordem da ordem” são as próprias coisas, a realidade, os objetos, as peças de museus, enfim as coisas “reificadas” tomadas no pólo de sua materialidade e, podemos acrescentar, o conhecimento sobre a realidade de mundo (a realidade espiritual contida nos livros). Já a “segunda ordem da ordem” são as descrições, como os catálogos de fichas, existentes fisicamente e refletindo, por meio das classificações, a divisão e a hierarquia de seus objetos. Na terceira “ordem da ordem” todo o conteúdo pode ser

“digitalizado”, bem como sua descrição nos metadados, e estão disponíveis na forma de *bits e bytes* no ambiente digital, ou seja, não existe “uma ordem” única.

Monteiro (2003) relata que o paradigma da organização do conhecimento é, em última análise, o mesmo da linguagem, ou tecnologia da inteligência proeminente em determinada época, no caso, da linguagem verbal escrita. A escrita foi o primeiro sistema a tornar possível o conceito de “universal”, uma vez que o signo sedentário, permanente e fixo possibilita instaurar todas as práticas hermenêuticas e universais, e nesse contexto, aprimorando o paradigma da escrita, a organização do conhecimento buscou no significado uma classe ou um termo que vale como referente ontológico.

Para representar as coisas com clareza, precisamos organizá-las e para organizá-las precisamos de identidades fixas e definidas. É nessa seara que o logocentrismo reforça as verdades universais reduzindo a realidade à identidade do conceito. Essa é a crítica especial de Derrida (2000) e Deleuze e Guattari (1995, v.2) sobre o regime significante do signo imperante na modernidade.

Assim que as modalidades significativas aristotélicas, os predicáveis e as categorias, importantes na estrutura hierárquica do conhecimento, na sua classificação e indexação, assistem a “planificação” no mundo virtual. Nesse contexto, o léxico no ciberespaço não tem um termo, classe ou mesmo um significante adotado como “certo” à organização e busca do conhecimento e, portanto, não há uma “sintaxe geral” e universal. Com efeito, a quebra da sintaxe ou a multiplicidade é característica do pós-moderno: cada mecanismo tem a sua ou múltiplas sintaxes.

Em outra oportunidade já associamos o hipertexto e os mecanismos de busca aos princípios do rizoma, de Deleuze e Guattari (1995, v.1), a saber: conexão, heterogeneidade, multiplicidade, a-significante e cartografia (MONTEIRO, 2008).

A multiplicidade semiótica das formas simbólicas no ciberespaço e a desterritorialização dos signos, que o virtual permite, explicam a indexação baseada em múltiplas sintaxes. A partir dessa premissa, da qual estamos investigando, vamos apresentar um panorama dos mecanismos de busca no ciberespaço no tocante à proposição de uma tipologia.

Conceitos novos como, resultados orgânicos¹, resultados pagos, visibilidade das páginas e indexação (já conhecido) fazem parte dessa nova realidade e, de acordo com Fragoso (2007), na ausência de um controle por *gatekeeping* “na entrada”, o ambiente da *Web* propicia a emergência de mecanismos de filtragem e de seleção “na saída” do sistema.

Assim sendo, estabelecer um panorama dos mecanismos de busca, mesmo que provisório, dada a dinamicidade desse objeto, nos parece importante para conhecer essa realidade, pois a busca se tornou em um procedimento universalmente utilizado para buscar informação e conhecimento no ciberespaço.

2 TIPOLOGIA DOS MECANISMOS DE BUSCA

Os mecanismos de busca recebem várias nomenclaturas na literatura científica como, buscadores, ferramentas de busca, serviços de busca, motores de busca, entre outros que, doravante, serão denominados “mecanismos de busca”.

Dessa forma, a partir do Quadro 01, iremos apresentar uma tipologia e discuti-la e ilustrá-la a partir da tese das múltiplas sintaxes de organização e busca do conhecimento no ciberespaço, das quais incluem (mas não se esgotam) alguns exemplos infracitados:

CATEGORIAS	CLASSIFICAÇÃO	EXEMPLOS
ANATOMIA	<ul style="list-style-type: none"> • <i>Crawling</i> (varrer) • <i>Indexing</i> (indexar ou gerar o índice a partir da base de dados) • <i>Searching</i> (buscar através da interface de busca) 	
FORMA GERAL DE ORGANIZAÇÃO OU INDEXAÇÃO (<i>indexing</i>)	<ul style="list-style-type: none"> • Diretórios ou Catálogos • Programas ou robôs de Busca • Híbridos • Metabuscares 	<ul style="list-style-type: none"> ➤ Yahoo! ➤ Google ➤ Yahoo! ➤ Kartoo; Metacrawler
ORDENAÇÃO DOS RESULTADOS (<i>searching</i>)	<ul style="list-style-type: none"> • Localização/freqüência do termo • Análise de <i>links</i> • Relevância 	<ul style="list-style-type: none"> ➤ Google e todos ➤ Google (autoridade e eixo) ➤ Cuil

¹ São os advindos da busca algorítmica ou editorial.

APRESENTAÇÃO DOS RESULTADOS (<i>searching</i>)	<ul style="list-style-type: none"> • Agrupamento ou clusterização: <ul style="list-style-type: none"> ➤ Textual ➤ Visual • Especializados • Personalizados • Ontoweb • Web Semântica 	<ul style="list-style-type: none"> ➤ Clusty ➤ Mooter ➤ North light ➤ Google Acadêmico ➤ Buscapé ➤ Google Earth ➤ Google Blog Search ➤ Fizy, Justhearit ➤ Windows Live Search ➤ A9 ➤ Ontoweb
PARADIGMA SEMIÓTICO (<i>indexing e searching</i>)	<ul style="list-style-type: none"> • Sonoro • Visual • Textual • Híbridos 	<ul style="list-style-type: none"> ➤ Musipedia, Midomi ➤ Retrievr (Systemone); Google Earth ➤ Google ➤ Lygo, oSkope, kmap, Cuil

Quadro 1 – Tipologia dos mecanismos de busca

2.1 Anatomia

Essas máquinas de busca operam três processos principais no ciberespaço, a saber:

- a) *crawling*;
- b) *indexing*;
- c) *searching*.

Na primeira etapa um programa denominado *crawler*² “varre” o ciberespaço, periodicamente, com critérios específicos (algoritmos) coletando todas as páginas possíveis da *Web*. A partir da base de dados gerada pelo *crawler* gera-se o índice e por fim a interface de busca do usuário.

Desse modo, o *crawler* navega na *Web*, de conexão em conexão, colhendo as páginas que encontra e enviando-as para ser indexadas em sua base de dados. Ele também anota as conexões das páginas e envia novas solicitações às conexões recém-

² Matthey Gray criou o Wanderer (1993), isto é, programa de varredura do conteúdo no ciberespaço e uma interface de busca que permitia o usuário consultar o índice. Mais tarde, criou o *crawler* colocando-o sobre um algoritmo de largura para transpor muitos *sites* antes de se aprofundar na indexação. Um pouco mais tarde, Brian Pinkerton (1994) desenvolveu o *WebCrawler*, que foi o primeiro a indexar todo o texto dos documentos que encontrava na *Web* (BATTELLE, 2006).

encontradas, as quais acham mais conexões, e assim *ad infinitum*. Nessa etapa, do processo de organização do conhecimento, os mecanismos de busca usam vários robôs, ou *crawler* para cobrir grande extensão da *Web* e construir a base de dados.

O próximo passo é a geração do índice, que separa os dados em várias partes, fazendo listas de palavras associadas com *URLs* (*Uniform Resource Locator*). Cada mecanismo de busca tem seu algoritmo, pelo qual faz a análise de acordo com algum critério específico de indexação. Para Battelle (2006, p. 17):

[...] desde o final da década de 1990, o índice tornou-se uma importante área de inovação para todas as empresas de busca – onde é aplicada uma grande parte do molho secreto de uma ferramenta de busca. Pense no índice como uma enorme base de dados de informações importantes a respeito de sites da *Web*. Empresas inovadoras como o Google construíram sua reputação estudando essa base de dados – observando padrões estatísticos e os potenciais em termos de algoritmos, descobrindo novas maneiras de alavancá-la no sentido de meta suprema de fornecer resultados mais relevantes às suas perguntas.

Na análise, o índice “etiqueta” seus dados por meio dos metadados, ou seja, procede a indexação para fornecer os resultados de busca. De acordo com Cendón (2001), a maioria dos mecanismos de busca inclui cada palavra do texto visível das páginas, outros extraem apenas o URL e as palavras que ocorrem com frequência, ou palavras e frases mais importantes contidas no título ou nos cabeçalhos e nas primeiras linhas, outros extraem os termos das *metatags*, etiquetas fornecidas para classificação do *site* por meio do campo “descrição” ou “palavras-chave”.

A próxima etapa é a interface de busca para o usuário, pois os dados após analisados, indexados e etiquetados:

Eles são jogados naquele que é chamado um índice de tempo de execução (*runtime index*) – uma base de dados pronta para oferecer resultados aos usuários. **O índice de tempo de execução forma uma espécie de ponte entre a parte dos fundos (seu crawl e seu índice) e a parte da frente (seu servidor de perguntas e a interface com o usuário).** (BATTELLE, 2006, p. 18, grifo nosso).

Cendón (2001) chama essa etapa de “motor de busca propriamente dito”, onde a consulta é recebida e transmitida para o *software* de busca, e esse programa também é responsável pela ordenação dos resultados, de acordo com os critérios de cada mecanismo.

Apesar de a grande parte da inteligência de uma ferramenta de busca estar contida na análise, do ponto de vista de Battelle (2006), o servidor ou a interface também tem

recursos que garantem uma potencialização da sintaxe de busca, como por exemplo, *sites* que indagam “Você quis dizer...”. Além da pesquisa básica, que utiliza as estratégias booleanas para refinar o argumento de busca, tem-se a avançada que traz resultados por idiomas, extensão de arquivos, data, domínio, etc.

Em síntese, todo processo de pesquisa realizado por máquinas exige a captura, a indexação e a interface ou programas de busca.

A partir da compreensão do funcionamento básico dos mecanismos de busca iremos apresentar um panorama geral, que pretende, por meio de investigação, desenvolver uma tipologia dos vários buscadores, para estudar as múltiplas sintaxes de organização e busca do conhecimento no ciberespaço, baseada nos seguintes atributos:

- a) forma geral de organização ou indexação;**
- b) ordenação e apresentação dos resultados;**
- c) paradigma semiótico.**

2.2 Forma Geral de Organização ou Indexação

Os mecanismos dividem-se, classicamente, em diretórios ou catálogos, programas ou robôs de busca, sistemas híbridos e metabuscadores. Os robôs, utilizados pelos programas, são ferramentas que “varrem” o ciberespaço para coletar informações para a geração do índice em sua base de dados, “[...] também são chamados de aranhas (*spiders*), agentes viajantes (*wanderers*), rastejadores (*crawles*) ou vermes (*worms*) [...]”. (CENDÓN, 2001, p. 41).

Na literatura é comum a categorização dos “tipos” de mecanismos que denominamos, neste estudo, de forma de geral indexação, por organizar o conhecimento, seja por categorias, como nos diretórios ou catálogos, como o Yahoo!, seja pela sintaxe apresentada nos textos, como nos mecanismos em que os robôs indexam o conteúdo sem uma categoria estabelecida, como no Google.

Essas categorias temáticas são classificadas hierarquicamente, conforme a sintaxe da lógica aristotélica, do gênero (geral) para o específico (espécie). O Yahoo!³ ilustra um

³ De acordo com Battelle (2006) o nome Yahoo! tem duplo sentido: significa rude e não sofisticado e também a palavra se prestava à engenharia reversa por meio da sigla “Yet another Hierarchical Officious Oracle” (Mais um Oráculo Hierárquico Oficioso).

mecanismo de busca organizado por diretório, e seus resultados são apresentados, em um primeiro momento, por *sites*.

Quando da criação do Yahoo!, uma abordagem hierárquica de um diretório fazia mais sentido, frente ao caos do início da *Web*, do que uma caixa de busca que demanda conhecimento de refinamento dos argumentos de busca, pois “As pessoas eram apanhadas no debate diretório *versus* busca.” (YANG *apud* BATTELLE, 2006, p. 51). Não obstante, com a popularização dos mecanismos “**A busca começou a fazer mais sentido como metáfora de navegação.**” (BATTELLE, 2006, p.51, grifo nosso).

O Yahoo!, atualmente, é considerado híbrido, pela possibilidade de se buscar um assunto diretamente pela caixa de entrada (CENDÓN, 2001), ou mesmo pela possibilidade de buscar as páginas da *Web*. Isso quer dizer que, hoje, quase não há mais distinção entre os diretórios e os programas de busca, conforme Figura abaixo.

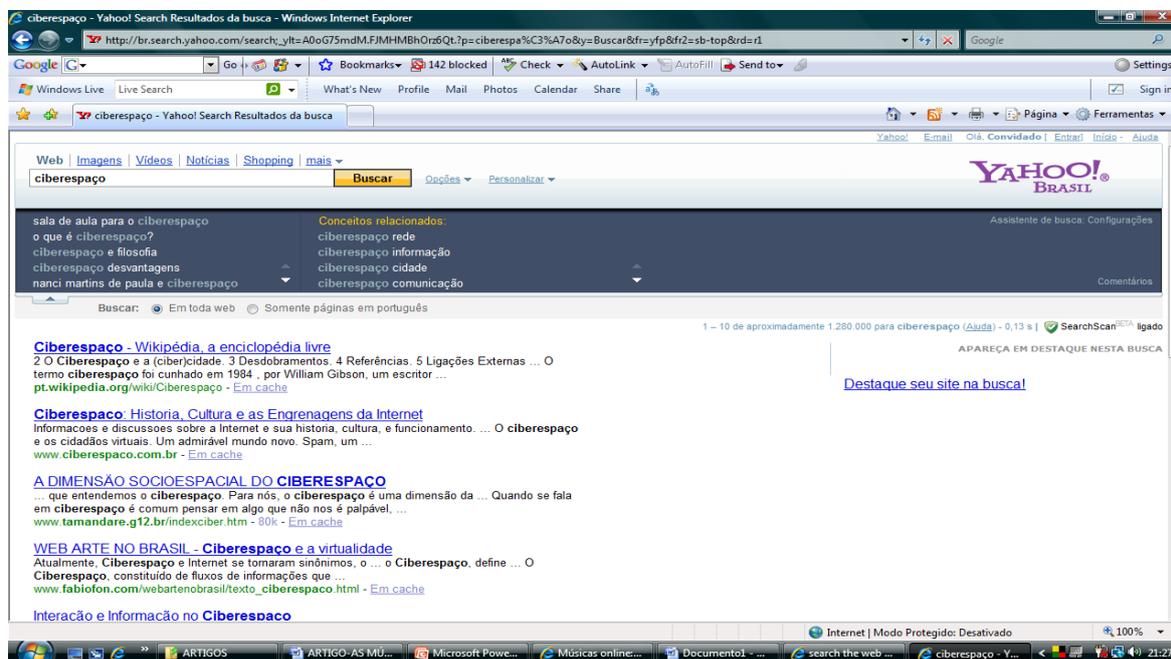


Figura 1 - Busca na *Web* pelo Yahoo!

Nessa direção, muitos mecanismos também oferecem forma de inclusão de *sites* enviada pelo autor para uma indexação mais rápida, não sendo, portanto, uma atribuição particular dos diretórios. Por outro lado, o Google, buscador eminentemente indicial acrescentou a possibilidade de busca por categorias, conforme a Figura 02:

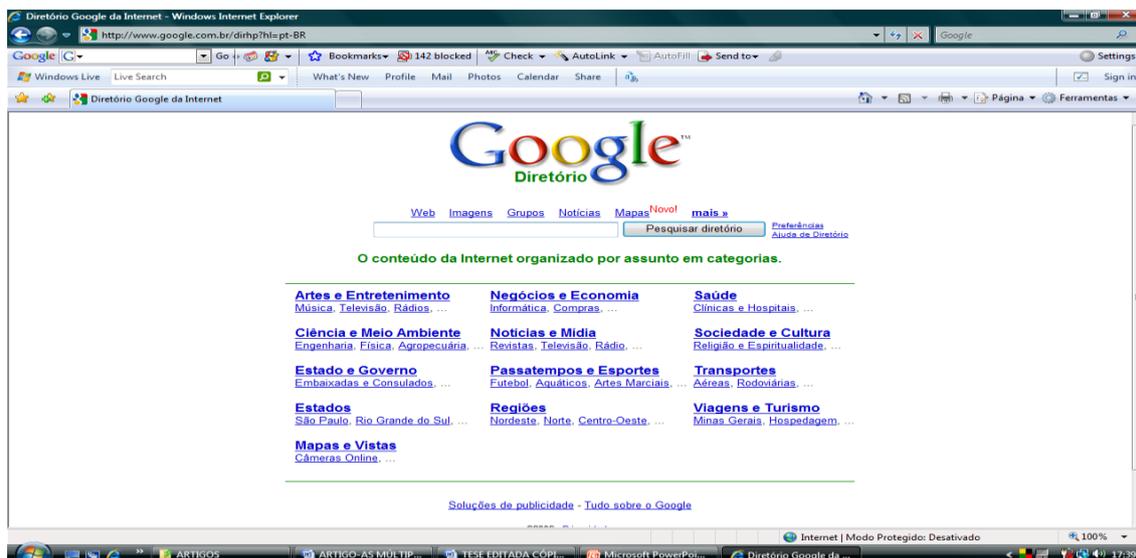


Figura 2 - Diretório Google da Internet

Isso reforça a tese que “Diretórios” e “Programas” estão se mesclando em uma terceira tipologia: os mecanismos “híbridos”.

Já os metabuscadores, sejam textuais ou visuais, fazem suas buscas em bases de dados dos outros mecanismos, não realizando a indexação propriamente dita, sendo também, ao nosso ver, um tipo de fonte por agrupamento.

O Metacrawler (Figura 03) é um metabuscador que agrupa resultados de vários buscadores, como Google, Yahoo!, Ask About (entre outros) dos quais são informados na descrição dos resultados “*found on...*” possibilitando também, a busca de imagens, vídeos, notícias, entre outros.

O Metacrawler, quando surgiu, resolveu as diferenças de sintaxe entre os vários buscadores consultados, criando a sua própria sintaxe e convertendo [...] o *input* do usuário no comando correspondente em cada sistema de busca acessado.”(FRAGOSO, 2007, n.p.).

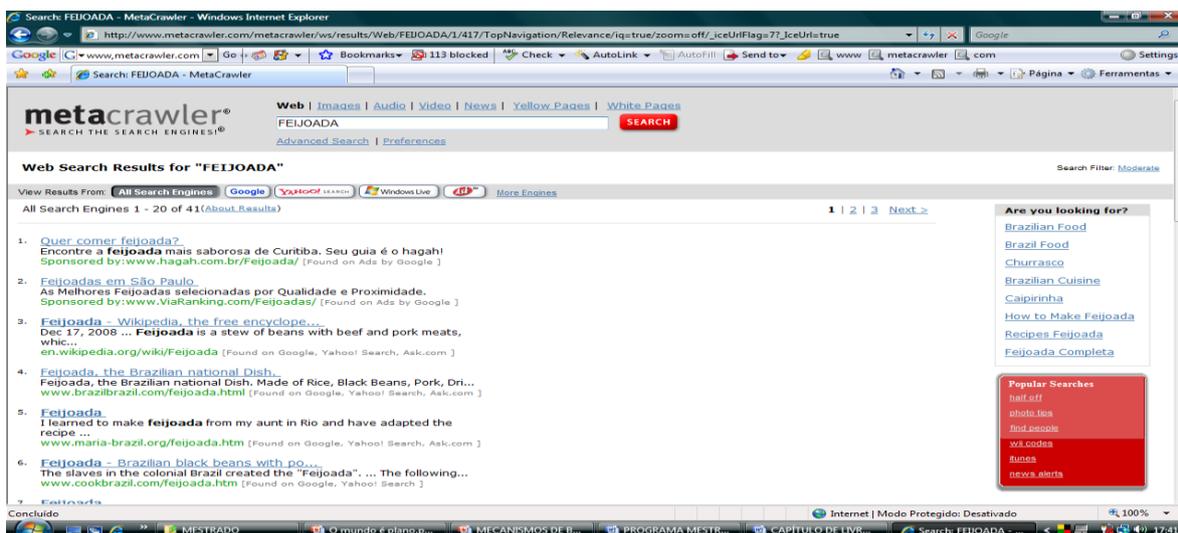


Figura 3 - Resultado de busca do Metacrawler

O KartOO é um metamotor que representa seus resultados por mapas interativos, em que, esferas maiores ou menores (passando o cursor) significam a pertinência com o assunto buscado, conforme Figura 04. Esse buscador é mais interessante que o Mooter (mais à frente), tanto visualmente, quanto em termos de apresentação dos resultados, trazendo, ainda, resultados de imagens e vídeos sobre o assunto buscado.

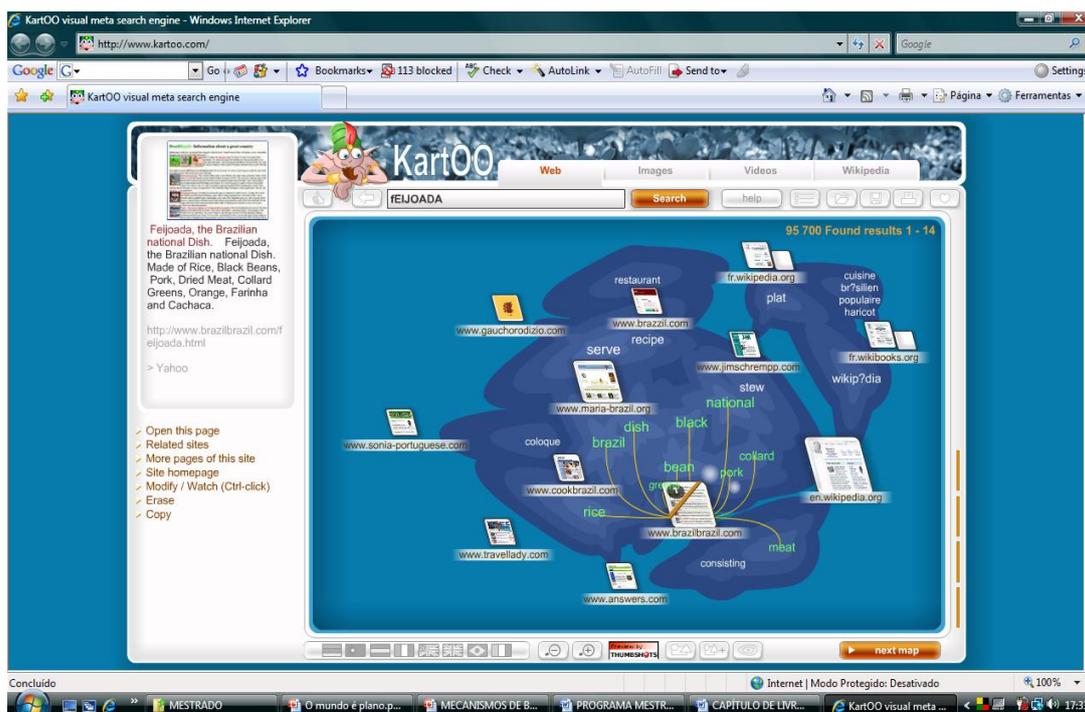


Figura 4 - Mapas interativos do KartOO

2.3 Ordenação dos Resultados

De acordo com o Quadro 01, iremos falar sobre alguns critérios de ordenação de resultados, que segundo Battelle (2006) é a tecnologia que faz a ligação com o índice e a interface de busca. Existem outros algoritmos e tecnologias para estabelecer critérios de ordenação, mas não são descritos na literatura, por razões óbvias de sigilo.

Nessa direção, Fragoso (2007) nos chama a atenção que os buscadores funcionam como verdadeiros *gatekeepers* digitais, no entanto, operam conforme critérios mantidos em sigilo e objetivos comerciais. Enfatiza ainda que é sabido que: as ferramentas de busca tendem a indexar mais *sites* dos EUA, misturam resultados orgânicos e pagos, seus algoritmos podem ser manipulados interna ou externamente, entre outros.

Localização/ Frequência dos Termos

As palavras semanticamente significativas para a busca são analisadas de acordo com a frequência e localização das mesmas. Para Mostafa (2005) esse é um método de ordenação comum nos mecanismos de buscas e, hoje, está associado com outros critérios, como o Google, por exemplo, que analisa mais de 100 fatores para determinar a relevância de um *site* para as suas palavras-chave. Detalhando melhor esse critério, Cendón (2001) explica que existem buscadores que analisam o termo no título, em cabeçalhos de destaque ou nos primeiros parágrafos do texto.

Outros consideram o número de termos presentes em uma página e a proximidade entre eles. Outros levam em consideração o critério de densidade, ou seja, comparando o tamanho dos documentos *versus* a frequência do termo, os documentos mais curtos seriam mais “densos” em relação ao tema. Há ainda o critério da “curva declinante”, em que a primeira ocorrência do termo é mais relevante que a segunda, a terceira, e assim sucessivamente. Já o critério da “relação inversa”, isto é, palavras que recebem menos peso em razão de sua maior frequência, pois possuem pouco valor semântico, se aplica às expressões como: o, e, ou, ser (*stop words*).

Análise de Links

Significa a associação com outras páginas e sua importância ou relevância. Baseados nas análises de citações, os então acadêmicos Larry Page e Sergey Brin

(Google) começaram a observar os resultados de pesquisa na *Web* e perceberam que os *links* apontavam para as *webpages* para frente, mas não registravam as conexões para trás, ou seja, os “*back links*” que cada *site* recebia.

Ademais, pensaram que poderiam levantar a importância de quem se conectava a quem, pelo “impacto” das citações, seja pela autoridade ou pelo eixo. Assim, foi criado o *Pagerank*, um algoritmo que conta, tanto o número de conexões em determinado *site* quanto o número de conexões em cada um dos *sites* interligados, da seguinte maneira:

- autoridade: muitas outras páginas apontam para ela, isto é, ela é muito citada;
- eixo: ela aponta para muitas outras, ou seja, ela cita muitas páginas.

Uma página pode ter alto *Pagerank* se houver muitas páginas apontando para ela, ou se houver algumas páginas de alto *Pagerank* apontando para ela. Intuitivamente, páginas que são frequentemente citadas em muitos outros pontos da web são merecedoras de uma visita. (PAGE; BRIN, 1998, p. 110).

Para Diaz-Isenrath (2005) esse critério recursivo de relevância poderia ser comparado, na verdade, a uma capacidade de voto, quer dizer, quanto maior o *Pagerank* de uma página, mais *Pagerank* ela pode passar para outras.

Além da análise de *links*, o Google possui ainda, como exemplo de múltiplas sintaxes, a opção “mais”, em que agrega outros buscadores especializados em algum tipo de busca como, mapas, livros, imagens, diretórios, textos acadêmicos, entre outros, e aplicativos inteligentes para refinar a busca, como “você quis dizer...” ou mesmo na caixa de busca a ocorrência de termos mais procurados a partir do argumento de busca digitado.

Relevância

De acordo com o *site* Cuil (Figura 05), mais que simples métrica, o buscador oferece seus resultados de acordo com a relevância do conteúdo de suas páginas, analisando a palavra-chave com o conteúdo global, conceitos e a coerência das páginas inter-relacionadas nos resultados.

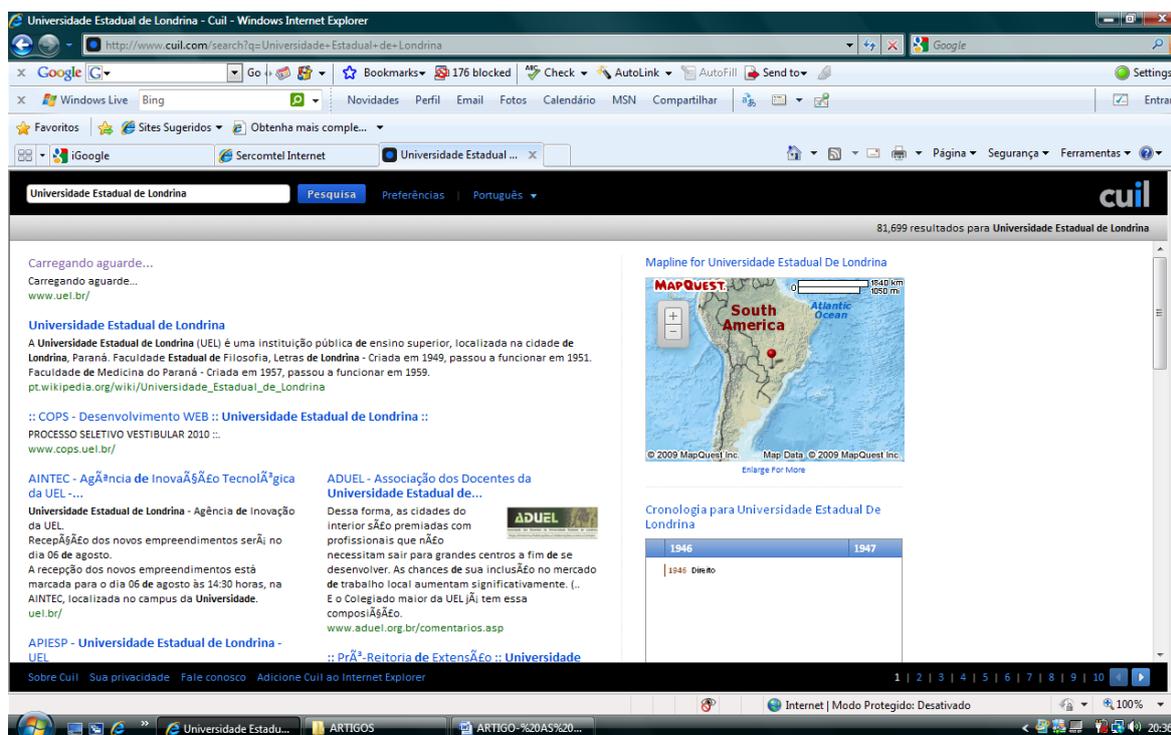


Figura 5 - Elementos relevantes no resultados de busca no Cuiil

Para Fioravante (2009, p. 01) a relevância surge como principal vetor de novas tendências dos mecanismos de busca, uma vez que a fase de mapeamento de conteúdo, no ciberespaço, já é passado, “Assim, a relevância de uma busca só existirá se a ferramenta de busca conseguir, de forma eficaz, identificar o contexto da pesquisa e seus elementos.”

Como cada algoritmo tem a sua maneira (sigilosa) de “ranquear” o conteúdo, assim, pode-se resumir a relevância dos documentos conforme sua relação espacial das palavras-chave utilizadas com o texto, para a busca (FRAGOSO, 2007) e, sobretudo, o contexto dos resultados com o argumento de busca utilizado, diferentemente de resultados listados sob a forma de índice.

2.4 Apresentação dos Resultados

A apresentação dos resultados de alguns mecanismos será visualizada para ilustrar as múltiplas sintaxes de busca de informação e conhecimento no ciberespaço. Pela extensão do artigo, apenas alguns foram selecionados para discussão.

Há uma relação precípua entre o ordenamento e a apresentação dos resultados e segundo Batelle (2006), a interface de busca é a ferramenta que garante a ligação da

“parte dos fundos” com a “parte da frente” dos mecanismos de busca. Na Filosofia da linguagem, lembramos do conceito deleuzinano “a dobra” em que sentido permanece na fronteira entre as coisas e a linguagem, é a articulação entre homem, máquina e linguagem, na superfície. A dobra “[...] é a continuidade do avesso e do direito, a arte de instaurar essa continuidade, de tal maneira que o sentido na superfície se distribui dos dois lados ao mesmo tempo.” (DELEUZE, p. 130, 1998).

Agrupamento ou Clusterização

Como mencionado no Quadro 01, supracitado, há os mecanismos de busca que trazem os resultados de acordo com algum agrupamento lógico, seja verbal ou visual, em que selecionamos alguns exemplos para ilustração.

Desses, o melhor é o Clusty, em que o agrupamento é claro e informa, a partir de coluna apresentada na página de resultados, o tipo de “clusterização”: todos os resultados, tipos de fontes e extensão das URLs.

No primeiro caso, “todos os resultados”, o Clusty apresenta um agrupamento da ocorrência do léxico mais recorrente nos resultados, conforme se pode observar a partir da busca por “feijoada”.

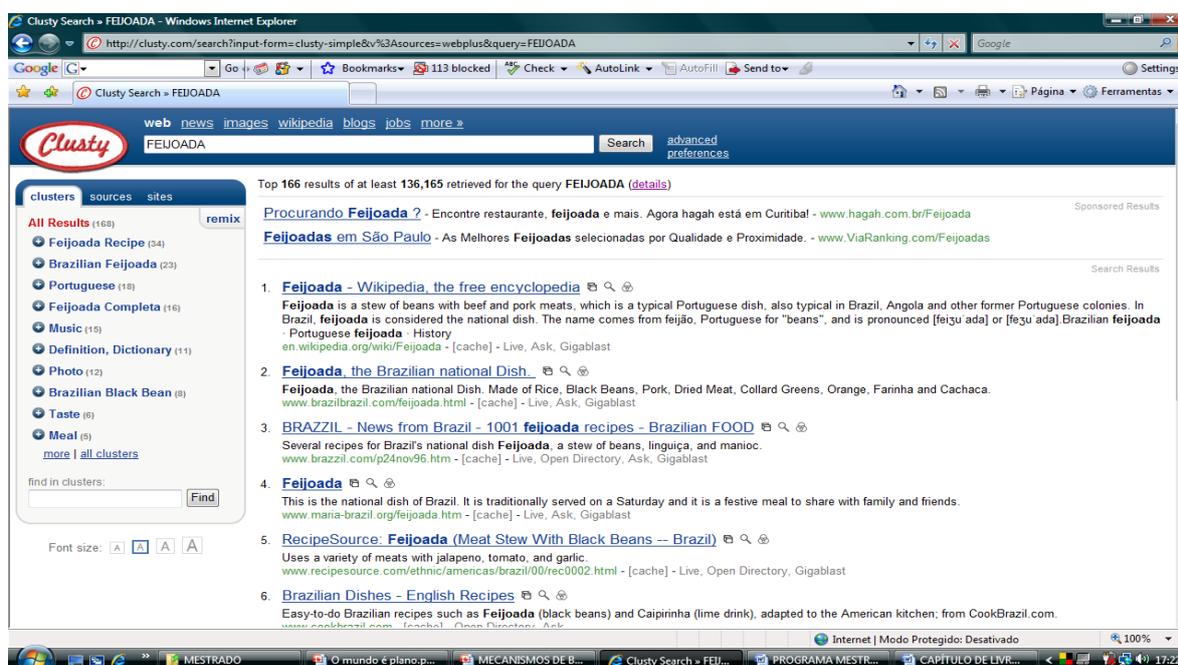


Figura 6 - Resultado de busca no Clusty

A segunda guia de agrupamento, da Figura 07, apresenta as fontes, ou seja, os sites e buscadores que foram consultados para retornar os resultados da pesquisa.

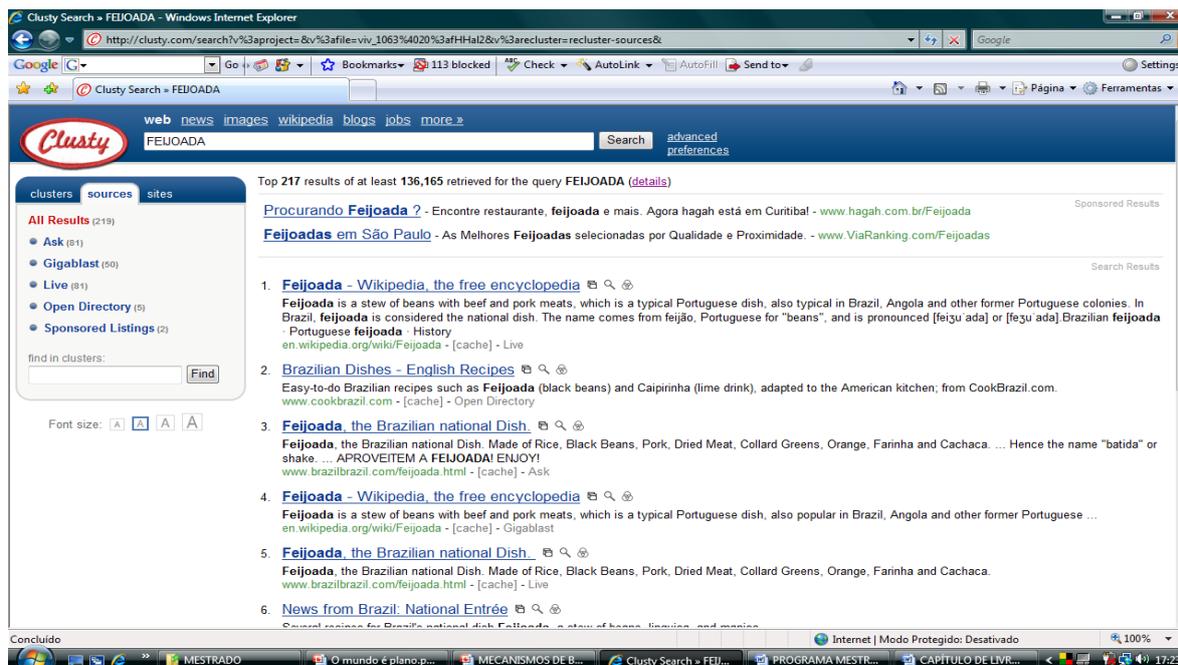


Figura 7 - Resultado de busca no Clusty de acordo com seleção de tipo de fontes

Já a figura abaixo apresenta a extensão dos sites recuperados a partir da busca, ou seja, comerciais, organizacionais e, outras URLs.

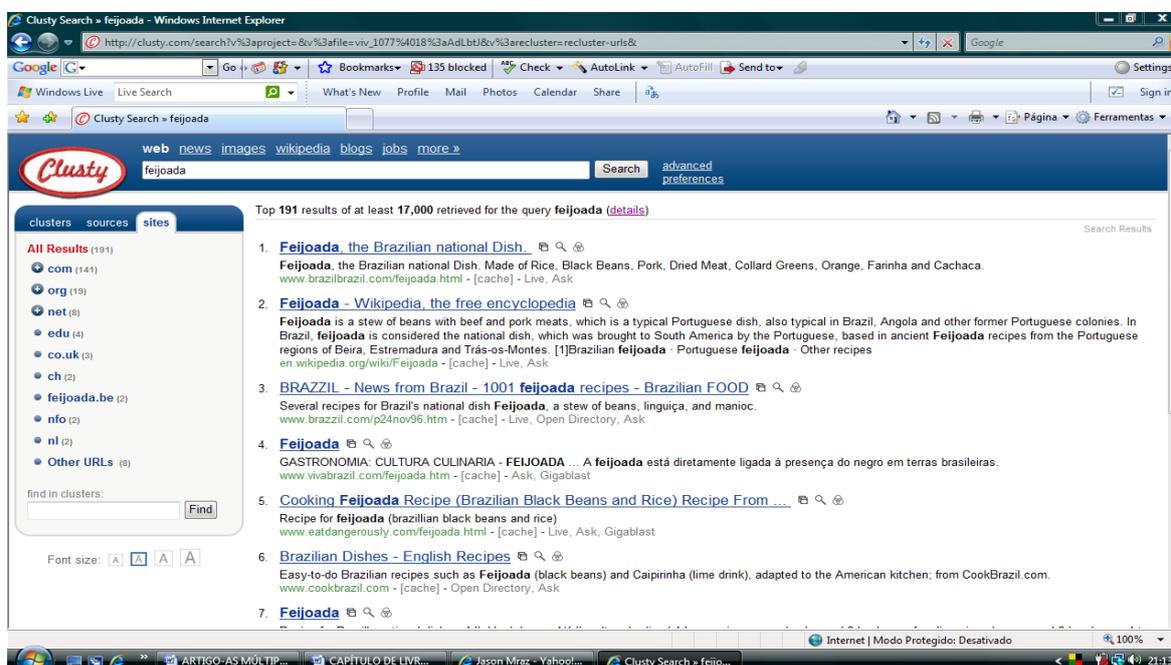


Figura 8 - Resultado de busca no Clusty por extensão da URL

Já o buscador Mooter (Figura 09) apresenta seus resultados de acordo com agrupamento visual, por meio de botões, da ocorrência de alguns léxicos no argumento de busca “feijoada”. Apesar de se autodenominar com o “poder da relevância”, em algumas buscas tal agrupamento não parece ser tão relevante ou lógico, como no Clusty. Observa-se também que, o buscador não é totalmente visual, uma vez que opera a intersemiose, isso é, a palavra, no ato na busca, bem como nos resultados verbais, diferentemente da classificação semiótica apresentada mais à frente.

Não obstante, em alguns assuntos, o Mooter agrupa os botões em subcategorias de forma mais lógica e traz ainda a opção “refinar”, memorizando os grupos escolhidos, que combina a consulta atual com grupos consultados anteriormente, fornecendo resultados mais precisos (MOSTAFA, 2005).

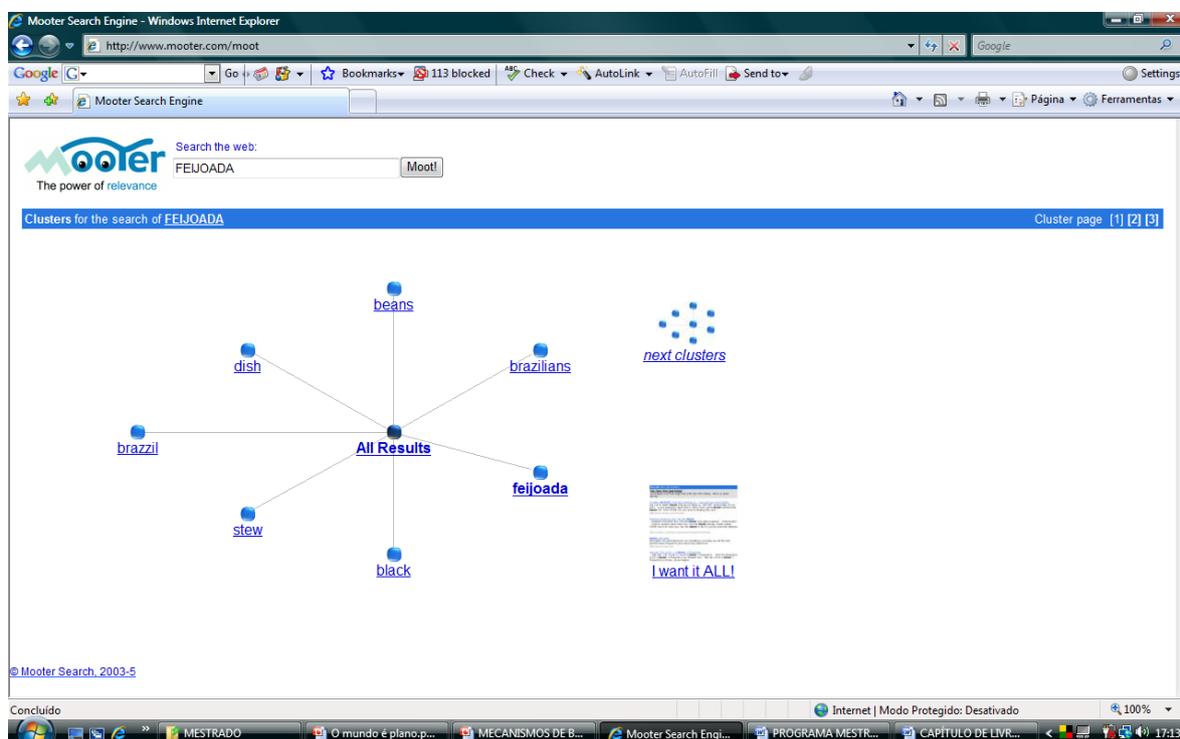


Figura 9 - Agrupamento visual no Mooter

Especializados

Para essa tipologia procurou-se agrupar mecanismos de busca especializados em algum tipo de atividade, seja acadêmico, comercial, por tipo de comunicação, como os buscadores de blogs, e assim por diante. Mais uma vez percebemos as múltiplas sintaxes de indexação, em que várias atividades e formas simbólicas e semióticas são indexadas

no ciberespaço, não ficando circunscritas somente no âmbito do discurso científico, como na ordem bibliográfica tradicional.

O Google é um exemplo de múltiplas sintaxes, seja em seu buscador original, seja no avanço que faz no sentido de diversificar suas operações de busca no ciberespaço, indo em direção à ubiquidade (digitalizando e indexando conteúdo analógico e múltiplos setores), conceito necessário para construir uma verdadeira memória digital e, conseqüentemente, a busca perfeita, que consiste em conhecer o comportamento de busca e as preferências do usuário. Com efeito, “O Google é, atualmente, a maior declaração do poder da busca de nossa cultura [...]” (BATELLE, 2006, p. 3143).

Para ilustrar esta seção, destacamos o Google Acadêmico.

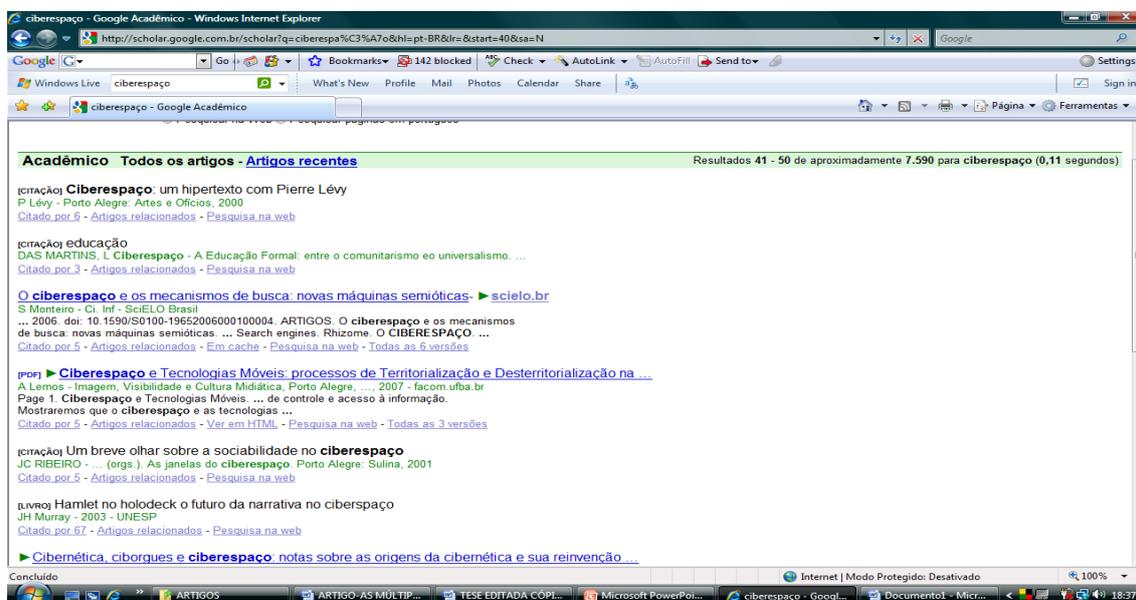


Figura 10 - Google Acadêmico

Segundo o *site* da empresa, o Google Acadêmico indexa artigos revisados pelos pares (*peer-review*) e publicações de cunho acadêmico. Classifica os resultados de pesquisa segundo a relevância, onde as referências mais úteis são exibidas no começo da página. A tecnologia de classificação do Google leva em conta o texto integral de cada artigo, o autor, a publicação em que o artigo saiu e a frequência com que foi citado em outras publicações acadêmicas. Vale observar, nos resultados, os seguintes campos:

- citado por: identifica quantos e quais documentos citaram o texto;
- artigos relacionados: seleciona outros artigos relacionados semanticamente ao tema;

- c) pesquisa na *Web*: o buscador posiciona o resultado do texto na *Web*;
- d) *link* para biblioteca: logo após o título do artigo, uma flecha indica o repositório ou a biblioteca que contém os documentos;
- e) versões: apresenta os vários repositórios que contêm o artigo.

Entre outras novidades, o Google Acadêmico oferece suporte para que as bibliotecas possam, por meio do “resolvedor de *links*”, mostrar seus resultados nesse buscador.

No próximo exemplo ilustramos um mecanismo de busca especializado em música, em que se pode ouvir e assistir vídeos sobre determinado cantor.

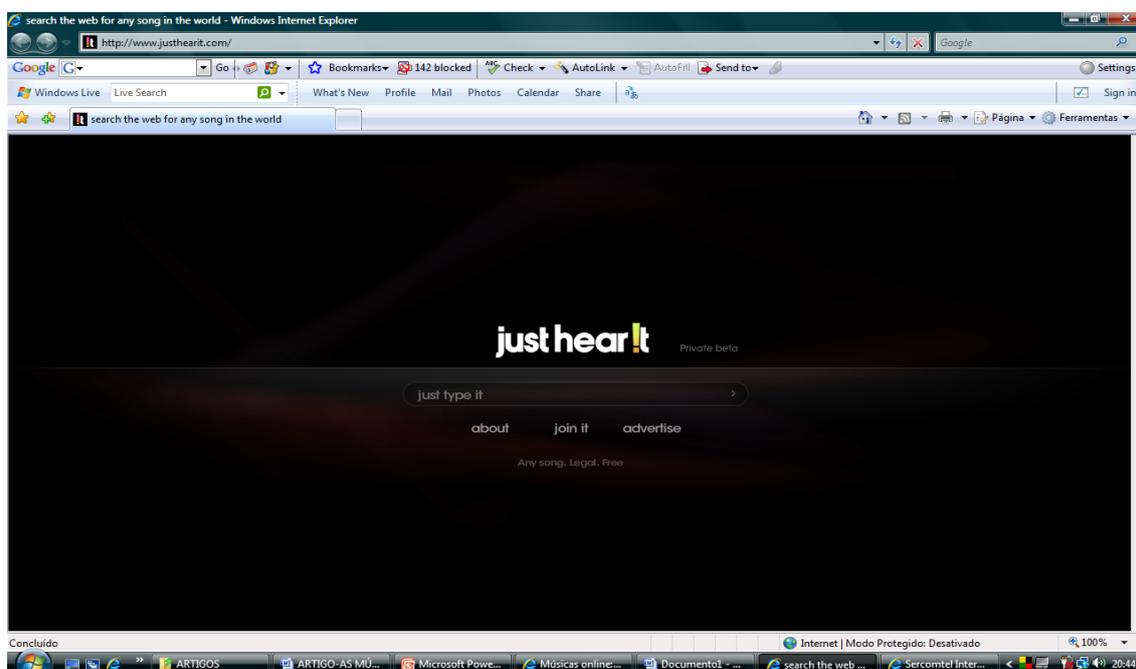


Figura 11 - Buscador especializado em música Justhearit

Esse buscador se diferencia dos listados como “sonoros” em nossa classificação semiótica, uma vez que a consulta é feita pela palavra para buscar música, operando, portanto, a intersemiose, de modo que sua classificação como especializado em música, nesta seção, é mais apropriada.

Já buscadores especializados em cotações e compras possuem uma tecnologia capaz de indexar o conteúdo dinâmico do ciberespaço, como banco de dados das empresas, para a consulta e busca dos resultados, citando como exemplo o *site* Buscapé.

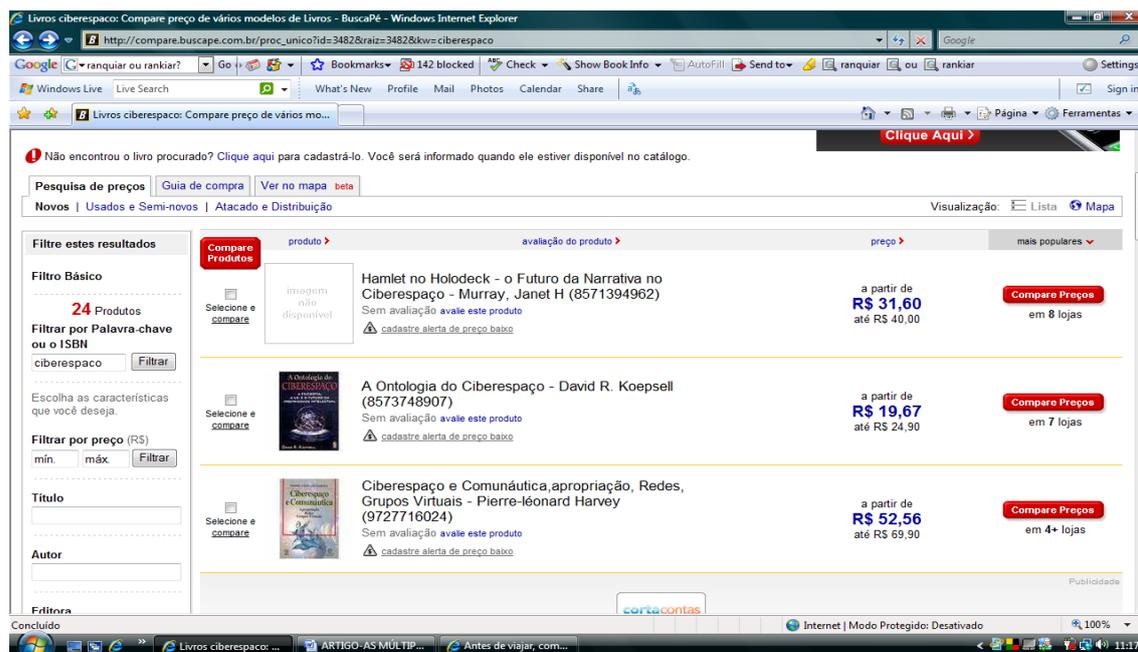


FIGURA 12 - Busca por produto no site Buscapé

Assim, alguns buscadores estão desenvolvendo um tipo de *software* denominado *wrapper*, que “[...] exploram a sintaxe habitual dos pedidos de busca e os formatos padrão dos recursos *on-line* para obter acesso a conteúdos ocultos.” E são denominados, por Mostafa, (2005, p. 87) de “mecanismos superiores”.

Personalizados

Alguns *sites* comerciais já apresentam páginas personalizadas com o nome do usuário, histórico de consulta e compras, como o Submarino e o Amazon, por exemplo, entretanto, para Mostafa (2005, p. 88):

Todos esses recursos novos, porém, são melhorias graduais. Mas, para determinar o contexto do usuário, os projetistas de *software* precisam superar sérios obstáculos de engenharia. Os desenvolvedores primeiro precisam criar sistemas que monitorem automaticamente os interesses e hábitos de um usuário, de modo que os mecanismos de busca possam determinar o contexto em que a pessoa está conduzindo uma busca de informações.

Com relação aos buscadores, o A9, Ask Jeeves e o Google criaram recursos para os usuários personalizarem as buscas.

Dentre aqueles buscadores que permitem personalizar as buscas, ilustramos o Windows Live Search que apresenta ao usuário uma interface de busca “customizável” que permite personalizar a busca de acordo com alguns atributos, inclusive algum controle sobre o “ranqueamento” dos resultados: mais populares, mais bem classificados e mais recentes.

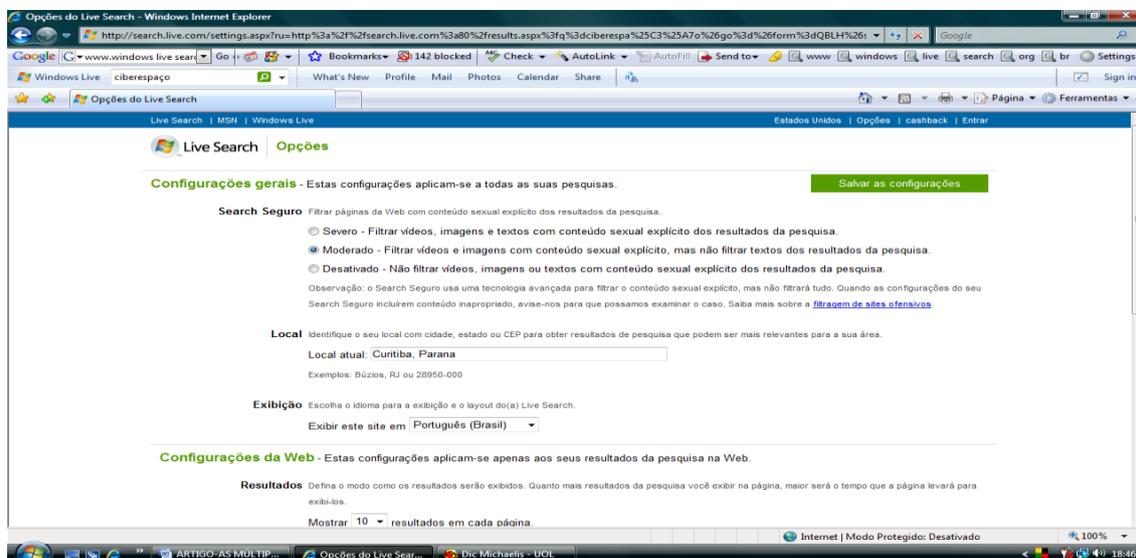


Figura 13 - Personalização do Windows Live Search

Dessa forma, o Live Search propicia configurações gerais para todas as pesquisas de acordo com a filtragem escolhida, localidade, idioma para exibição do *layout* do site, apresentação dos resultados e idiomas selecionados para busca (estes dois últimos, comuns em vários mecanismos).

O Google experimentalmente colocou “no ar” a possibilidade de personalizar a ordenação dos resultados, embutindo um programa wiki para tecer comentários dos resultados, conforme Figura abaixo:

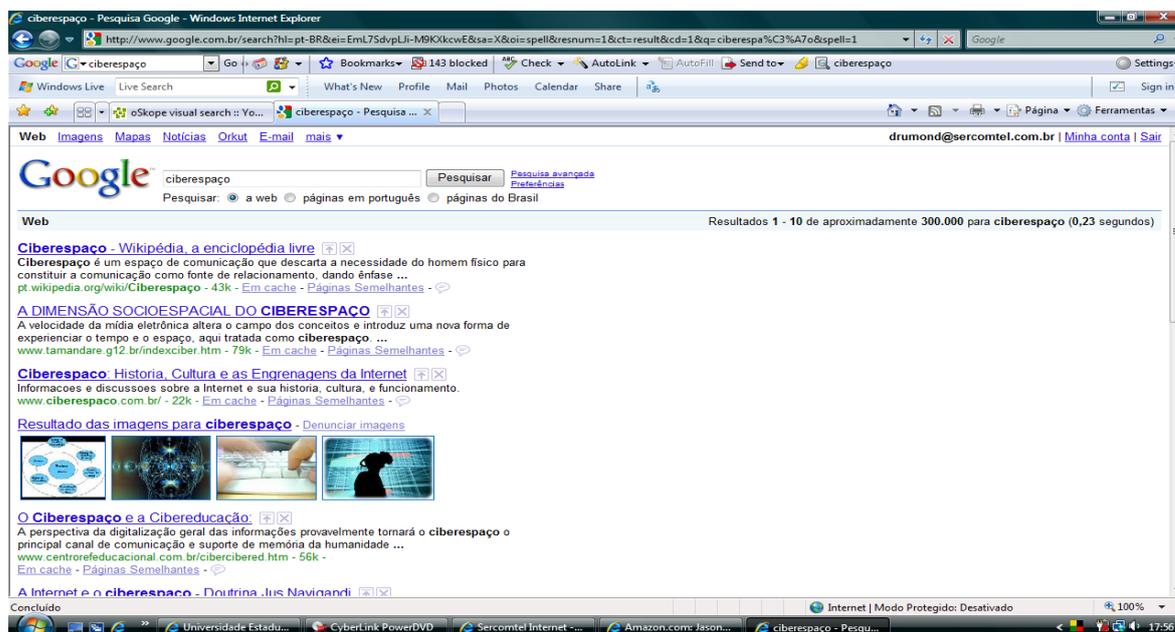


Figura 14 – Personalização da ordenação dos resultados no Google

Ontoweb

De acordo com Moraes, Ambrósio e Gomes (2005, p.1) a associação de semântica às páginas da *Web* vem sendo feita, portanto, por meio do uso de metadados e ontologias, pois “Na prática, ontologias servem para definir categorias para as coisas que existem em um mesmo domínio, além de definir regras de combinação entre seus termos e relacionamentos.”

Como exemplo, o “OntoWeb” está focado no cenário de **Governo Eletrônico**, sendo uma ferramenta informacional especialmente dirigida a esse contexto, tanto no que diz respeito às fontes de informação, quanto a sua estrutura de ontologias. Com algoritmos potentes e validados por processos científicos, o OntoWeb contempla uma nova modalidade de estruturas informacionais e representação do conhecimento, proporcionando buscas mais precisas e detalhadas. O **ONTOWEB®** está registrado no INPI, e consolida as bases conceituais da tecnologia KMAI® e de outras tecnologias e metodologias essencialmente brasileiras e ratificadas pela comunidade científica internacional, as quais, somadas a técnicas de Inteligência Artificial, geram importantes diferenciais, como a possibilidade de utilização de mais de 15.000 caracteres para análise: o campo de pesquisa não é limitado a palavras-chave ou a simples expressões de busca, de acordo com a Figura 15.

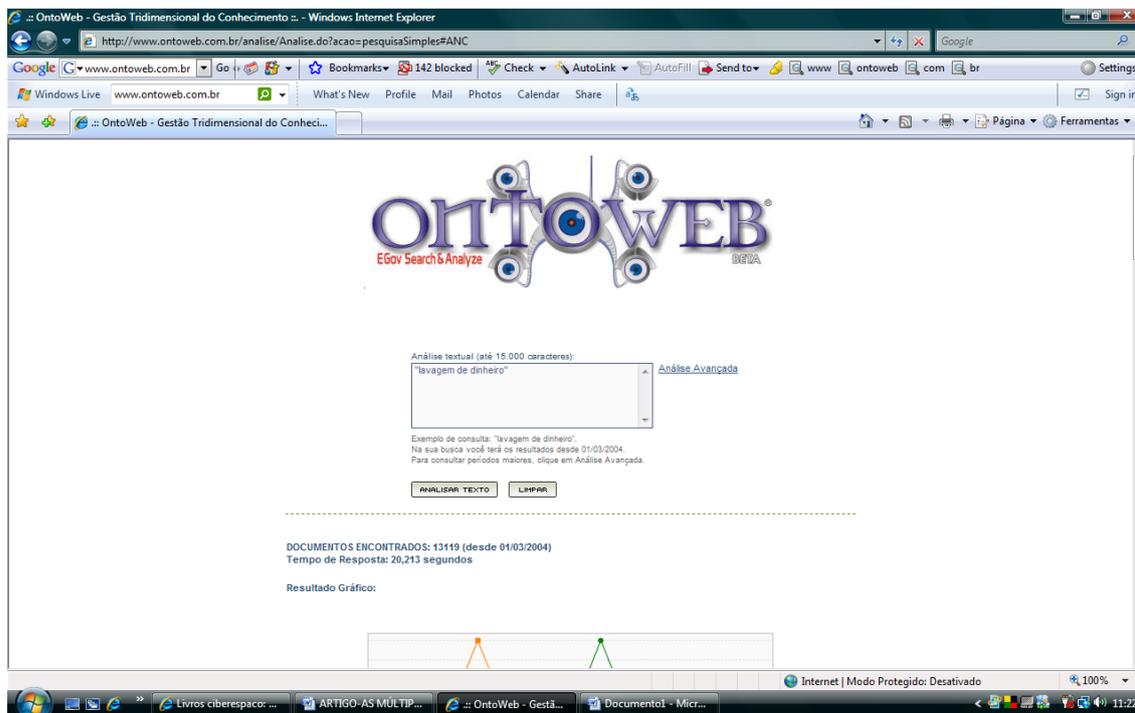


Figura 15 – Campo de busca do Ontoweb

Apresenta também gráfico de linhas na resposta: por meio dele é possível acompanhar a variação do assunto pesquisado no tempo, gerando subsídios para uma análise qualitativa mais eficiente, conforme Figura 16.

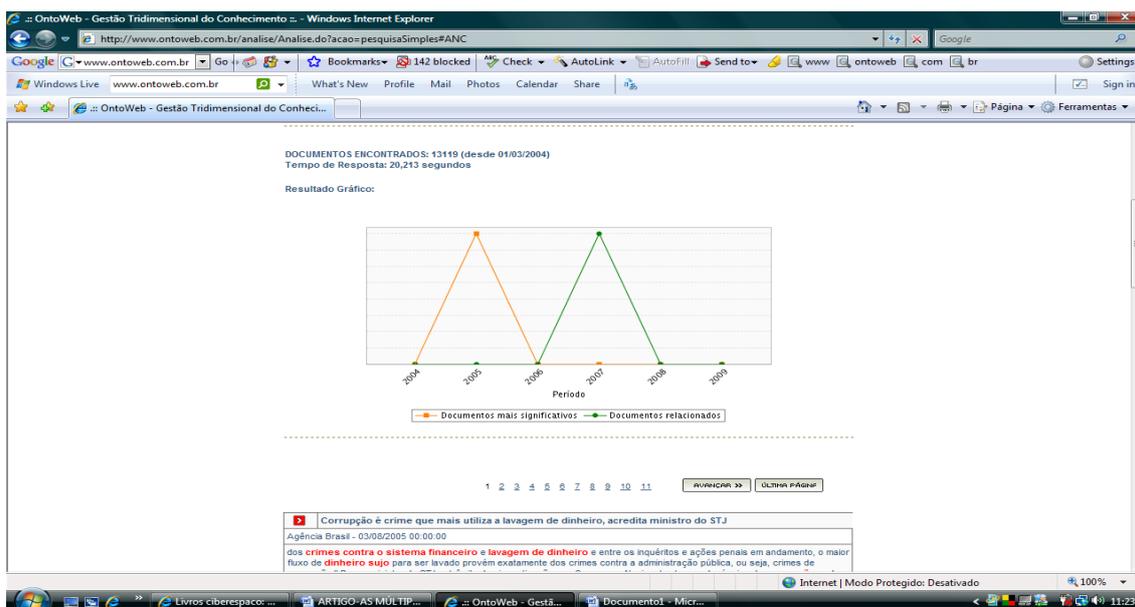


Figura 16 – Resultado gráfico da busca no Ontoweb

O Ontoweb utiliza ontologias contextualizadas no processo de busca: expandindo os conceitos utilizados na pesquisa e identificando seu contexto é possível localizar os documentos mais adequados à demanda solicitada (ONTOWEB, 2008).

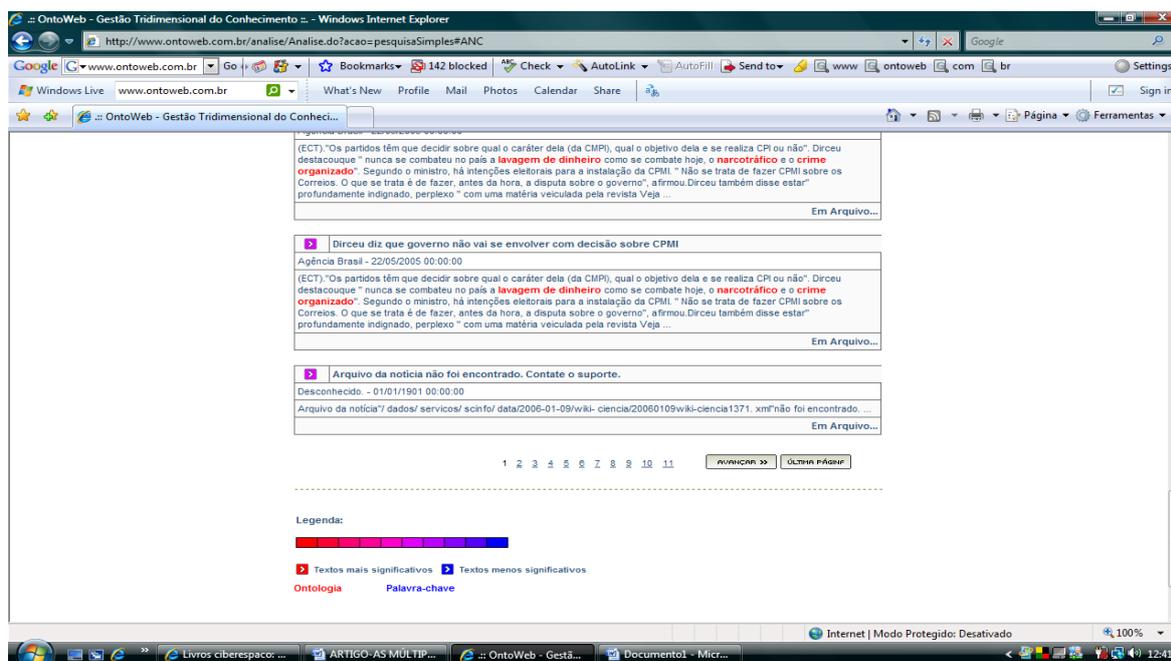


Figura 17 – Resultado de busca no Ontoweb

2.5 Paradigma Semiótico

Fixamos como eixo de estudo o paradigma semiótico, por entender que são tecnologias e paradigmas diferenciados dos buscadores anteriores arrolados. Esta categorização, para a pesquisa que estamos desenvolvendo, é de extrema importância, pois reflete o momento em que o conhecimento está rompendo com a cultura verbalista e o ciberespaço desterritorializando os signos, permitindo todas as hibridizações possíveis. Signos sensíveis e inteligíveis, misturados na construção de novas formas simbólicas, e os mecanismos de busca seguindo a mesma rota de organização.

Assim, não devemos confundir uma pesquisa de imagem em um buscador verbal, em que a palavra está no lugar lógico de todos os metadados, como argumento de busca, seja pelo título, pelas *tags*, descritores da imagem, seja pelo título de uma canção para buscar uma música. Nesses casos, ocorre ainda a intersemiose, com a proeminência do código verbal, ou seja, uma matriz semiótica (a palavra) sendo usada para recuperar outras linguagens.

Nesta tipologia, pensamos na forma análoga de busca, ou seja, usar o som para buscar o som, a forma para imagem, diretrizes espaciais para localização geográfica e a palavra para os textos.

Para tanto nos apropriamos das teorias de Peirce para compreender as matrizes semióticas dos buscadores. Para Peirce, a fenomenologia tem por função responder algumas questões: como se dá a apreensão e compreensão do mundo pelo homem? Como a multiplicidade e diversidade infinitas do universo sensível são convertidas em realidade inteligível? (*apud* SANTAELLA, 2001).

Assim, depois de décadas de reflexão, Peirce concluiu que tudo o que aparece à consciência, assim o faz numa gradação de três elementos: 1) qualidade ou sentimento (primeiridade); 2) reação (secundidade); 3) mediação (terceiridade).

A partir das três categorias fenomenológicas propostas por Peirce, pode-se estudar ou mesmo classificar os eventos semióticos. Dessa forma, Santaella (2001) em sua teoria defende que a variedade e a multiplicidade de todas as formas de linguagem estão alicerçadas em apenas três matrizes, a saber: sonora (ou virtual), visual e verbal.

A sonora está contemplada no primeiro nível de semiose (primeiridade), ou seja, possui as características do quali-signo, ou qualidade de sentimento. Na primeiridade, ou primeira categoria fenomenológica da linguagem-pensamento, as coisas nos aparecem de forma imediata, original, livre e espontânea, a primeira forma de predicação das coisas (SANTAELLA, 1983). Aqui é a ambiência da iconicidade que, além de representar por qualidades de sentimentos, representa também por **percepções auditivas**. Queremos dizer que nem todo ícone é imagem. Aliás, as imagens são hipóícones.

Como exemplo de buscador sonoro, ilustramos o Musepedia, abaixo:

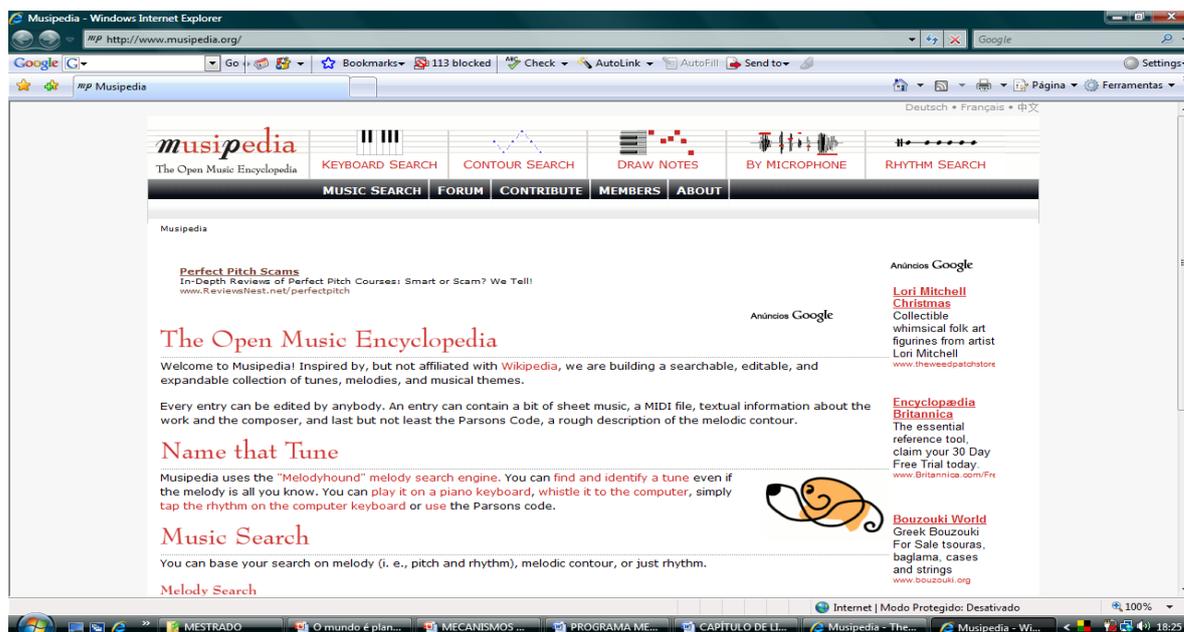


Figura 18 – Busca no Musipedia

Do ponto de vista da Semiótica, a linguagem sonora é a mais universal que existe. Talvez os programas que “baixam” as músicas e MP3 possam comprovar a grande audiência desse tipo de linguagem no ciberespaço. Traduzir para programas a qualidade de reconhecer sons para buscas, obviamente passa por outras linguagens computacionais mais complexas do que essa classificação inicial.

Quando se utiliza o signo sonoro ocupando o lugar lógico do som para a busca, percebe-se toda a dimensão do atributo utilizado por Santaella (2001) para classificar a linguagem sonora: A SINTAXE. Para proceder a pesquisa, nesse tipo de buscador, há de se ter conhecimento musical, isto é, dar a “partida perfeita” para buscar o pretendido. Isso requer noções de ritmo, harmonia ou notas musicais, a não ser que se insira a gravação da música como argumento de busca, que é a partida perfeita para tal ação (veja a barra superior do Musipedia).

Nesse caso, há a percepção como o signo verbal traduz todos os outros com tanta naturalidade que nem percebemos. Usar um mecanismo especializado em música utilizando a palavra é muito mais fácil, semioticamente falando, posto que incorporamos, por completo, a linguagem verbal escrita. Isso porque a palavra, pelo poder de tradução, pode estar no lugar lógico de todas as outras linguagens.

Evidencia-se também que as linguagens, de acordo com categorias fenomenológicas, são constituídas com sintaxes próprias. Isso significa que o hábito (no sentido peirciano) delimita o universo de percepção da linguagem para cada usuário.

A linguagem visual, de acordo com Santaella (2001), ocupa o segundo nível, ou a secundidade. Na secundidade, existir é estar em uma relação, e na classificação de Santaella (2001) os objetos são representados por conexão e referência. Com efeito, para Deleuze e Guattari (1995, v.2) os índices são os signos mais territorializados, dada à relação com a realidade.

Apesar das imagens representarem por semelhança da forma (isomorfismo, atributos do ícone), nesta classificação, e justamente por isso na classificação das linguagens, elas funcionariam como índices, ou seja, apontam uma relação ou conexão, por meio da forma, e não qualidades de sentimento. Assim também parece ser a lógica dos programas que fazem a busca baseados na relação da forma ou semelhança para localização da imagem buscada.

Essa classificação é possível por causa da proeminência dos signos, pois:

[...] o mundo não consiste de duas espécies exclusivas de coisas, signos, de um lado, e não signos, de outro, também não há três tipos mutuamente exclusivos de signos: ícones, índices e símbolos. Estes são, antes, elementos ou **aspectos de semiose** que, de uma semiose a outra, variam grandemente, na importância ou **proeminência** relativa. Podemos, para simplificar, chamar um signo pelo nome de **elemento ou aspecto que é mais preponderante nele, ou para o qual queremos chamar atenção**, o que não implica que não haja os elementos ou aspectos dos outros dois tipos. (FISCH *apud* SANTAELLA, 1992, p.84, grifo nosso).

Para ilustrar um mecanismo eminentemente visual apresentamos o Retrievr, do labs.systemone, conforme Figura 19:

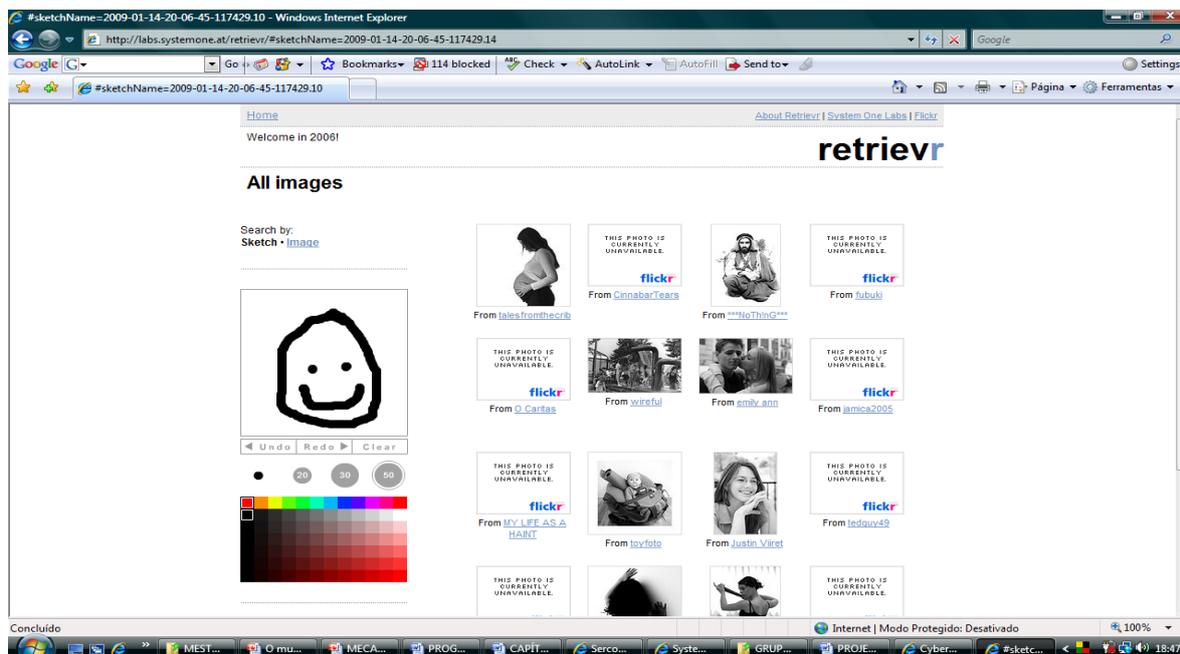


Figura 19 – Busca no Retrievr

No intento de buscar uma forma semelhante o mecanismo supracitado não funcionou muito bem, retornando apenas uma fotografia de uma moça sorrindo e talvez a forma arredondada de uma gravidez. Percebe-se, mais uma vez que, para o resultado ser satisfatório, há de se oferecer a partida ou argumento de busca perfeito, como no exemplo da Figura 20, em que o mecanismo oferece a importação da imagem para buscar outras correlatas na forma.

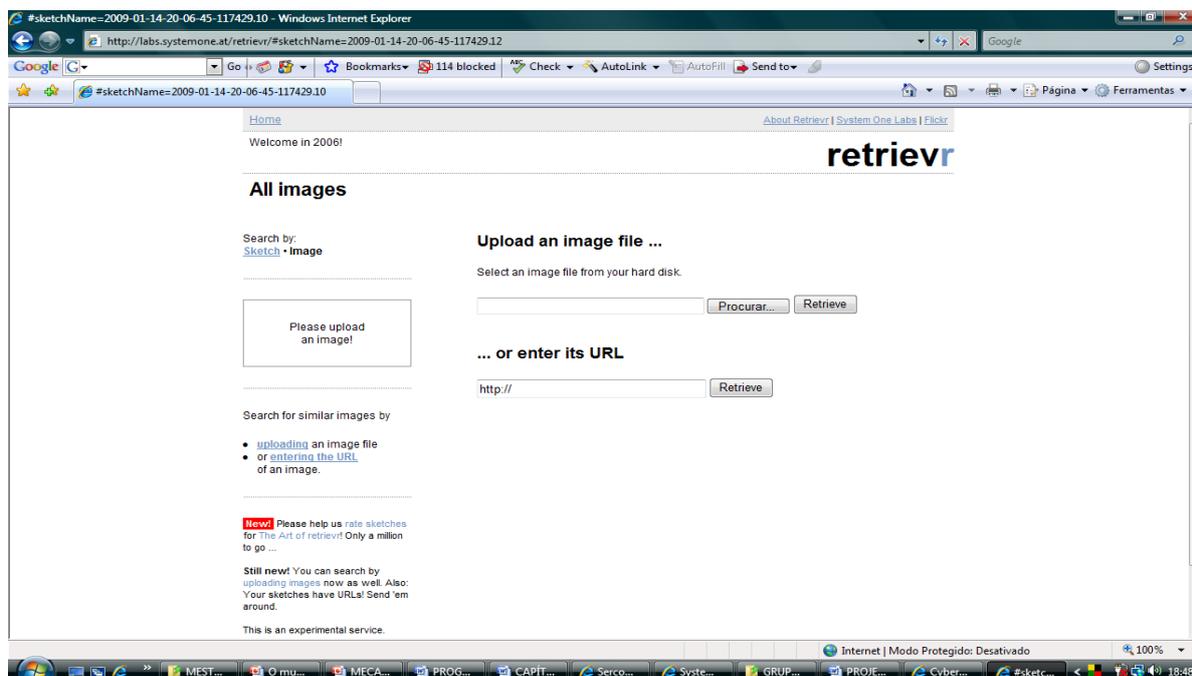
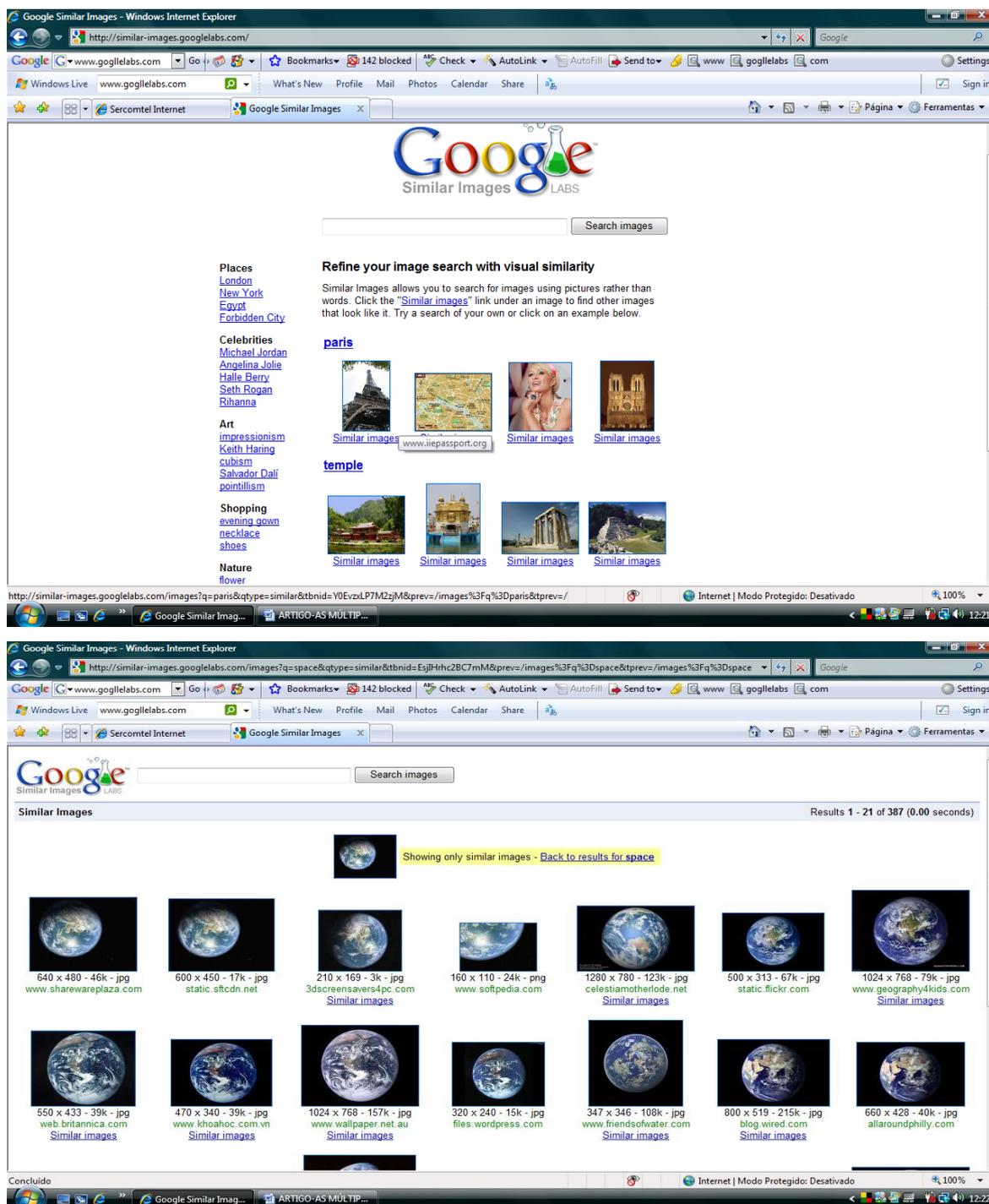


Figura 20 - Busca no Retrievr por importação de imagem

Como o futuro dos mecanismos de busca está na agregação semântica dos resultados, por meio de agentes inteligentes, outras plataformas não convencionais (dados móveis automotivos, entretenimento) e sistemas não-verbais, Mostafa (2005) acredita que os analistas do setor aprimorarão os recursos de extração de dados, no sentido de oferecer serviços de busca sofisticados integrados organicamente às funções visuais interativas, como no exemplo do buscador experimental do Google, constante nas Figuras 21 e 22:



Figuras 21 e 22 – Busca no Google Similar Images

A linguagem verbal escrita está no território da terceiridade, quer seja, da semiose, aprendizagem, cognição e lei, uma vez que é o campo da mediação do símbolo, da tradução e da reflexão.

Esses são os mais fáceis de utilizar, mesmo os mecanismos que operam por intersemiose (ou híbridos), pois a palavra tem esse poder de “tradução” de todas as outras linguagens posto a sua representatividade abstrata.

Cada uma dessas linguagens tem seu eixo de proeminência em uma das categorias, de modo que a linguagem sonora é uma questão de iconicidade ou qualidade, para quem ouve, pois para quem busca algum item, deve ter conhecimento ou hábito da sintaxe musical. A linguagem visual é uma questão de referencialidade, porém deve-se colocar a partida ou argumento perfeito em sua semelhança à recuperação satisfatória e a linguagem verbal traduz a representatividade de todas as outras, sendo a forma mais fácil de argumento de busca, seja em qualquer tipo de buscador, inclusive nos híbridos, devido ao seu poder de tradução.

Vemos proliferar uma gama de mecanismos híbridos, que misturam várias linguagens nos resultados de busca, dos quais podemos citar alguns exemplos muito interessantes: liveplasma.com; oskope.com; lygo.com; mnemo.org; kwmap.net, entre outros.

Esses, além da intersemiose entre o argumento de busca (palavra ou termo), os resultados estão imbricados, hibridados e entrelaçados numa miríade de signos especialmente imagéticos e verbais, tornando a busca mais interessante e intuitiva. Como exemplo, desta categoria, ilustramos o Lygos (Figura 23):

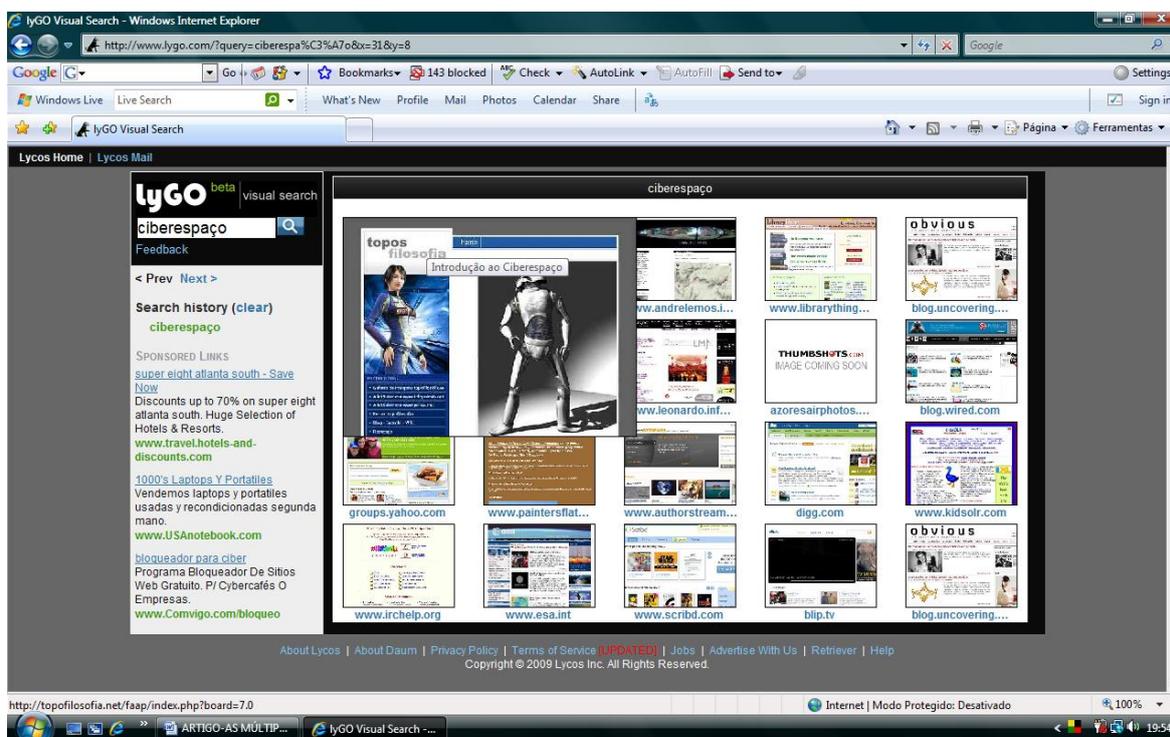


Figura 23 – Busca no Lygo

A hibridização é um fenômeno eminentemente rico semioticamente, tanto para Peirce como para McLuhan. Para este último autor, toda linguagem e meio estão ligados à percepção, os meios híbridos, porque misturam duas ou mais linguagens e/ou mídias, acabam também por estimular ou promover uma mudança na posição relativa dos sentidos (sensórios). Assim, o híbrido constitui um momento de verdade e revelação, do qual nasce a forma nova. “O momento de encontro dos meios é um momento de liberdade e liberação do entorpecimento e do transe que eles impõem aos nossos sentidos.” (McLUHAN, 1998, p. 75).

O ciberespaço é *per se* um ambiente híbrido, um espaço em que as três matrizes da linguagem-pensamento estão explicitadas em constante interação.

3 CONSIDERAÇÕES FINAIS

Se o projeto da modernidade realizou-se com o avanço das ciências e com as intervenções das classificações para a organização do conhecimento, que potencializa o sentido único dos referentes ontológicos, no pós-moderno, as Tecnologias da Informação e Comunicação vêm dando nova concepção ao conhecimento e a sua organização.

Para Weinberger (2007, p. 91-92), na modernidade, “[...] o conhecimento é um território que pode ser subjugado por meio da administração de uma metodologia rigorosa e inflexível [de classificação].” Já no ciberespaço:

A terceira ordem da ordem pega o território subjugado por classificação e o liberta. Em vez de impor-lhe categorias, a terceira ordem fixa *etiquetas* que permitem a um usuário de recursos *on-line* – páginas da Web, fotos – acrescentar uma palavra ou duas, de modo que possa localizá-los novamente mais tarde.

No ciberespaço as máquinas indexam os textos, não mais verbais escritos, mas híbridos, não mais fixos, antes, dinâmicos e desterritorializados, operando essa indexação na equivocidade e na polissemia da linguagem natural. Percebe-se a passagem do significado ou conceito adotado para os múltiplos sentidos e múltiplas sintaxes.

Os mecanismos atuais, mais populares no ciberespaço, compõem uma classe de sistemas de busca que se baseiam no código verbal – paradigma verbal escrito. São denominados sistemas de busca textual, embora alguns possam combinar em seus mecanismos mais de um paradigma para programar e operacionalizar a busca no ciberespaço. Mas, outros paradigmas semióticos são a base para novas classes de busca, embora o hábito ou conhecimento dessa sintaxe delimite, para a busca, o universo de percepção de cada linguagem para o usuário.

Conceitos como *Web* sintática, *Web* invisível e *Web* 2.0, com sua *folksonomia* e *tagosfera*, mais que separações, refletem um mar de significados e construções simbólicas que se misturam, atualizam-se, ou escondem-se na pragmática de conteúdos organizados e buscados pelos mecanismos de busca no ciberespaço. Já a *Web* semântica, extraindo algumas ferramentas já embutidas nos mecanismos e aplicações isoladas, em alguns domínios de conhecimento nas ontologias, é ainda uma construção situada em uma *Web* reduzida e a *Web* mundial sempre terá e trará o problema da taxonomia do conhecimento e da multiplicidade dos signos, seja em sua representação ou organização.

AGRADECIMENTOS

Externo meus agradecimentos ao CNPq e à Fundação Araucária pelo apoio recebido para a realização da pesquisa, em desenvolvimento, cujo artigo faz parte.

REFERÊNCIAS

- BATTELLE, John. *A busca*. Campinas: Campus; Rio de Janeiro: Elsevier, 2006.
- BAUDRILLARD, Jean. *Simulacros e simulações*. Lisboa: Relógio D'Água, 1981.
- BUSCAPÉ. Disponível em: <<http://www.buscape.com.br>>. Acesso em: 25 fev. 2009.
- CENDÓN, Beatriz V. Ferramentas de busca na web. *Ci.Inf.*, Brasília, v. 30, n. 1, p. 39-49, jan./abr. 2001.
- CLUSTY. Disponível em: <<http://www.clusty.com/>>. Acesso em: 12 abr. 2009.
- COSTA, Samuel Fernandes Magalhães. *Matrix e sua filosofia pós-moderna*. Disponível em: <[http:// WWW.chamada.com.br/mensagens/matrix.html](http://WWW.chamada.com.br/mensagens/matrix.html)>. Acesso em: 30 out. 2007.
- CUIL. Disponível em: <<http://www.cuil.com/>>. Acesso em: 05 set. 2009.
- DELEUZE, Gilles. *Lógica do sentido*. São Paulo: Perspectivas, 1998.
- DELEUZE, Gilles; GUATTARI, Félix. *Mil platôs: capitalismo e esquizofrenia*. São Paulo: Ed. 34, 1995. v. 1-2.
- DELEUZE, Gilles; GUATTARI, Félix. *Mil platôs: capitalismo e esquizofrenia*. São Paulo: Ed. 34, 1998. v. 5.
- DERRIDA, Jacques. *Gramatologia*. 2. ed. São Paulo: Perspectivas, 2000.
- DIAZ-ISENRATH, Cecília. Um estudo sobre Google: questões para uma leitura micropolítica das tecnologias da informação. *Liinc em revista*, v.1, n.2, p. 1001-126, set. 2005.
- FIORAVANTE, Felipe. *Tendências emergentes em mecanismos de busca*. Disponível em: <www.terraforum.com.br/>. Acesso: 20 abr. 2009.
- GOOGLE. Disponível em: <<http://www.google.com.br>>. Acesso em: 05 fev. 2009.
- FRAGOSO, Suely. Quem procura acha? O impacto dos buscadores sobre o modelo distributivo da *Word Wide Web*. *Eptic: Revista de Economia Política de las Tecnologias da La Información y Comunicación*, v.9, n.3, n.p., sept/dic. 2007. Disponível em: <<http://www.eptic.com.br>>. Acesso em: 15 mar. 2009.
- JUSTHEARIT. Disponível em: <<http://www.justhearit.com/>>. Acesso em: 12 abr. 2009.
- KARTOO. Disponível em: <<http://www.kartoo.com/>>. Acesso em: 12 abr. 2009.
- LYGO. Disponível em: <<http://www.lygo.com>>. Acesso em: 12 abr. 2009.
- LYON, David. *Pós-modernidade*. São Paulo: Paulis, 1998.
- MCLUHAN, Marshall. *Os meios de comunicação como extensões do homem*. São Paulo: Cultrix, 1998.

- METACRAWLER. Disponível em: <<http://www.metacrawler.com/>>. Acesso em: 12 abr. 2009.
- MONTEIRO, Silvana Drumond. O ciberespaço: o termo, a definição e o conceito. *Datagramazero - Revista de Ciência da Informação*, v.8, n. 3, p. 1-18, jun.2007. Disponível em: <http://www.datagramazero.org.br/jun07/Art_03.htm>. Acesso em: 17 fev. 2008.
- MONTEIRO, Silvana Drumond. Os mecanismos de busca: à guisa de uma tipologia das múltiplas sintaxes. In: TOMAÉL, Maria Inês (Org.). *Fontes de informação na Internet*. Londrina: EDUEL, 2008. Cap.5, p.97-122.
- MONTEIRO, Silvana Drumond. A organização virtual do conhecimento no ciberespaço. *Datagramazero - Revista de Ciência da Informação*, v. 4, n. 6, p. 1-24, dez. 2003. Disponível em: <http://www.dgz.org.br/dez03/Art_05.htm>. Acesso em: 17 fev. 2008.
- MOOTER. Disponível em: <<http://www.mooter.com/>>. Acesso em: 12 mar. 2009.
- MORAIS, Edison A. M.; AMBRÓSIO, Ana Paula L.; GOMES, Pedro Henrique. Ontologias como fonte de indexação semântica na Internet. SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, 11., 2005, Poços de Caldas. *Anais...* Poços de Caldas: PUC, 5-7 dez. 2005.
- MOSTAFA, Javed. WWW procura indolor. *Scientific American Brasil*, p. 85-91, 2005. Disponível em: <www.sciam.com.br>. Acesso em: 28 out. 2006.
- MUSIPEDIA. Disponível em: <<http://www.musipedia.org>>. Acesso em: 10 jan. 2009.
- ONTOWEB. Disponível em: <<http://www.ontoweb.com.br>>. Acesso em: 18 fev. 2009.
- PAGE, Larry; BRIN, Sergey. The anatomy of large-scale hypertextual Web search engine. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 7th, Brisbane, 1998. *Proceedings...* Publicado em COMPUTER NETWORKS AND ISDN SYSTEMS, v.30, n.1-7, p. 107-117, April 1998. Disponível em: <<http://www.sciencedirect.com/science/journal/01697552>>. Acesso em: 25 fev. 2008.
- RETRIEVR. Disponível em:< <http://labs.systemone.at/retrievr/>>. Acesso em: 15 fev. 2009.
- SANTAELLA, Lúcia. *A assinatura das coisas: Peirce e a literatura*. Rio de Janeiro: Imago, 1992.
- SANTAELLA, Lúcia. *Matrizes da linguagem e pensamento*. São Paulo: Iluminuras, 2001.
- SANTAELLA, Lúcia. *O que é Semiótica*. São Paulo: Brasiliense, 1983.
- SANTOS, Jair Ferreira. *O que é pós-moderno*. São Paulo: Brasiliense, 2006.
- WEINBERGER, David. *A nova desordem digital*. Campinas: Elsevier: Campus, 2007.
- WINDOWS LIVE SEARCH. Disponível em: <<http://www.live.com>>. Acesso em: 05 fev. 2009.
- YAHOO! Disponível em: <<http://www.yahoo.com.br>>. Acesso em: 05 fev. 2009.

Title

The multiples syntaxes of the search engines on cyberspace

Abstract

On post-modernity the machines organizes knowledge by search engines. To discuss and to understand this reality better, the article proposes a typology to those search engines, aiming to illustrate the knowledge's organization and the search of multiple syntaxes on cyberspace. In this sense, it proposes the classification categories: anatomy, that are the basics processes of the search engine's work, crawling, indexing and searching; the general method of indexation or organization, which are the directories, programs, meta-searches and hybrids searchers; the ordinance and presentation of the results, which reflects the multiple syntaxes, and the semiotic paradigms, which illustrate the search engine's multiple semiotic. From this classification scheme the search engines are exemplified and discussed, enclosing this scenery of multiples indexation syntaxes and searching of information and knowledge on cyberspace and, in the same way, it builds an ontology for this object. The cyberspace will always bring and emerges the knowledge taxonomy problem and the multiplicity of signals, in its representation or organization.

Keywords

Information and Communication Technology. Search engines. Cyberspace. Syntaxes. Semiotics.

Título

Las multiples sintaxis de los mecanismos de búsqueda em el ciberespacio

Resumen

En el periodo post-moderno las máquinas están organizando el conocimiento por medio de los mecanismos de búsqueda. Para discutir y conocer mejor esa realidad, el artículo propone una tipología para los mecanismos de búsqueda, y ilustra las múltiples sintaxis de organización y búsqueda del conocimiento e información en el ciberespacio. De esa forma, presenta los ejes de clasificación: anatomía, que son los procesos básicos de funcionamiento de esos buscadores, o sea: *crawling*, *indexing* y *searching*; la forma general de indexación u organización, que son los directorios, programas, meetabuscadores y híbridos; la ordenación y presentación de los resultados que reflejan lãs múltiples sintaxis; y el paradigma semiótico, que ilustra la multiplicidad semiótica de los mecanismos de búsqueda. A partir de esa tipología, los buscadores son exemplificados y debatidos, abarcando ese escenario de múltiples sintaxis de indexación e búsqueda de la información y conocimiento en el ciberespacio y al mismo tiempo construyendo una ontología de esse objeto. El ciberespacio sempre tendrá y traerá el problema de la taxonomía del conocimiento y de la multiplicidad de los signos, sea en su representación u organización.

Palabras clave

Tecnología de la Información y la Comunicación. Organización del Conocimiento. Mecanismos de Búsqueda. Ciberespacio. Sintaxis. Semiotica.

Recebido em: 12.05.2009

Aceito em: 31.08.2009