

## De Paul Otlet à *Web Semântica*: aportes teóricos sobre a organização do conhecimento

**Januário Albino Nhacuongue**

*Universidade Federal de São Carlos – UFSCar, Email: januario80@gmail.com*

**Moisés Lima Dutra**

*Universidade Federal de Santa Catarina – UFSC, Email: moises.dutra@ufsc.br*

### RESUMO

Este trabalho consiste numa abordagem de caráter qualitativo, de natureza básica, e analisa a área interdisciplinar da organização do conhecimento com base no levantamento bibliográfico. Neste contexto, revisita alguns aportes teóricos desde os sistemas de classificação bibliográfica, passando pelo Tratado de Documentação de Paul Otlet e pela computação interativa de Engelbart até a *Web Semântica*. Ao mesmo tempo em que traça um olhar a respeito da semântica nos ambientes de produção, difusão e apropriação da informação, também questiona os modelos evidenciados na Ciência da Informação que constituem a base para a compreensão das questões inerentes à semântica na informação, enquanto parte do objeto de estudo do campo. Conclui que a Ciência da Informação desenvolveu-se a partir da natureza prática, e a ausência de modelos teóricos consolidados na área dificulta a percepção dos alcances da *Web Semântica* para o objeto “informação”, a partir do contexto interdisciplinar. A semântica está ligada à linguagem, à complexidade da comunicação humana e precede a sintaxe. Por isso, independentemente do tipo da semântica tratada na *Web*, a sua contribuição está na organização do conhecimento, pois qualquer esquema de organização depende da estrutura cognitiva que permite a sua criação e compreensão. Indubitavelmente, este enfoque vai contribuir nas discussões sobre a organização do conhecimento dentro da Ciência da Informação, resgatando a figura de Otlet que, às vezes, é ignorada na literatura. De igual modo, vai subsidiar os fundamentos sobre a *Web Semântica* a partir de um olhar específico da Ciência da Informação.

**Palavras-chave:** Ciência da Informação. Semântica. *Web Semântica*. Organização do conhecimento. Compartilhamento informacional.

---

### 1 INTRODUÇÃO

A área de organização do conhecimento é objeto de diferentes perspectivas, contextos e periodicidades de análise por campos análogos no concernente à finalidade do uso do conhecimento por domínios de aplicação das necessidades específicas. A maioria das suas abordagens resulta de relações interdisciplinares entre áreas como Lógica, Administração, Engenharia de Produção, Biologia, Psicologia, Biblioteconomia, Ciência da Informação, Documentação, Ciência da Computação e outras afins.

A semântica, por sua vez, desenvolve-se numa estrutura simbiótica e complacente com a organização do conhecimento, fundamentada pelos critérios de significação impostos pelas vicissitudes dos ambientes de produção, tratamento, difusão, apropriação e uso da informação para o conhecimento em ação, ou seja, conhecimento integrado ao contexto, à cultura e às necessidades de indivíduos, grupos ou sociedade.

A relação entre a semântica e a organização do conhecimento é antiga. Nasce da cooperação que norteia a atividade de produção do conhecimento. Uma cooperação que depende de níveis de integração de dados e de controle da linguagem de especialidade, do vocabulário e da terminologia em vários contextos comunicacionais.

O objetivo principal deste trabalho é abordar os aspectos da semântica na Linguística e na *Web*, estabelecendo uma relação com a Ciência da Informação. Assim, pretende-se evidenciar a dificuldade existente na área para a percepção da semântica com base em únicos modelos teóricos de abordagem. Igualmente, demonstrar que a semântica está relacionada à linguagem, por isso, na *Web* pode ser percebida a partir da natureza de cada linguagem do hipertexto.

Ainda dentro dos objetivos, procuramos ressaltar que o fundamento da *Web Semântica* se encontra enraizado na organização do conhecimento e não no próprio ambiente da *Web*. Por isso, enquanto esquema de organização do conhecimento para o compartilhamento, a sua abordagem é anterior<sup>1</sup> aos ambientes de linguagens híbridas em que ocorre a produção, difusão e uso da informação, e que caracterizam a *Web*. Por outras palavras, é anterior à tecnologia que propicia a representação baseada no hipertexto que congrega no mesmo espaço as linguagens sonora, visual e verbal. Daí a inclusão da classificação de Dewey, da documentação de Paul Otlet, do sistema integrado de Douglas Engelbart que posteriormente permitiu a introdução dos computadores na produção, organização e recuperação da informação, entre outras contribuições relacionadas à *Web Semântica* no contexto da Ciência da Informação.

Um dos argumentos que nos permite situar a *Web Semântica* na organização do conhecimento pode ser constatado a partir das considerações de Guimarães (2015, p.15) sobre as três dimensões da área (epistemológica, aplicada e cultural), envolvendo, entre diversos aspectos, “a promoção da interoperabilidade e de uma recuperação da informação cada vez mais rápida,

---

<sup>1</sup> Referimo-nos aos ideais da organização do conhecimento que prevalecem no paradigma da *Web Semântica* e não à própria terminologia, linguagens ou camadas específicas da sua arquitetura.

*Informação & Tecnologia (ITEC)*, Marília/João Pessoa, v.3, n.1, p.138-158, jan./jun. 2016.

eficaz e culturalmente significativa”. Portanto, uma premissa subjacente nas motivações do projeto da *Web Semântica*.

A ideia de destacar a contribuição de Otlet na organização do conhecimento e relacioná-la ao cenário da *Web* não é nova. Por exemplo, Santarem Segundo (2011) relaciona a constituição e os desdobramentos das redes sociais aos conceitos de Paul Otlet. Na sua dissertação de mestrado, Fontoura (2012) questiona a constituição do campo da Ciência da Informação no período pós-guerra com base em práticas e disciplinas que vinham sendo desenvolvidas há séculos. Neste cenário, destaca a documentação de Otlet que respondia aos novos desafios da organização da informação massiva, na sua maioria, resultante das invenções dos finais de 1800. Do mesmo modo, Wright (2014) aborda diversos feitos sociais de Otlet dentro do paradigma da universalização do conhecimento. Judge (2001), por sua vez, relaciona Otlet à origem do hipertexto, com base na União das Associações Internacionais cujos desdobramentos estavam centrados tanto na metodologia e organização do conhecimento como na transdisciplinaridade.

Entre outros, os trabalhos citados provam que a preocupação com a organização, difusão e uso compartilhado do conhecimento, ideia prevalente na *Web Semântica*, é um assunto antigo. O nosso trabalho, por sua vez, partilha do mesmo entendimento, com a particularidade de se estender para assuntos semânticos, questionando as formas como a Ciência da Informação visualiza as respectivas contribuições para o seu objeto.

## **2 A ORGANIZAÇÃO DO CONHECIMENTO: SUBSÍDIOS TEÓRICOS**

Na abordagem das questões inerentes à organização do conhecimento adotamos o método qualitativo de natureza básica pela riqueza de detalhes descritivos de fenômenos com base nas variáveis levantadas nos objetivos. O método foi complementado pela técnica da documentação indireta que consistiu na pesquisa bibliográfica de alguns dos principais autores que discutem a organização do conhecimento na Ciência da Informação. Também, por um lado, trouxemos as contribuições de autores que relacionam a figura de Otlet à *Web*, e por outro, evidenciamos autores que tratam a semântica no olhar interdisciplinar da Linguística.

### **2.1 O CONHECIMENTO COMPARTILHADO**

Conforme dito anteriormente, a preocupação em relação à organização do conhecimento

para o acesso universal é antiga. Porém, nem sempre este movimento para o compartilhamento da informação teve um percurso consensual e triunfante. Alguns fatos do passado permitem nos constatar que, em certos momentos, foi caracterizado por sinuosidades de cunho político, ostensivo e ambicioso. Por exemplo, quando o rei Ashurbanipal consolidou o Império Sumeriano no século VII a.C. abasteceu a biblioteca real com livros apreendidos dos adversários derrotados. O faraó egípcio Ptolomeu I, por sua vez, ordenou que o seu exército confiscasse todos os livros a bordo dos navios que entravam no Porto de Alexandria para a construção da famosa Biblioteca de Alexandria. Já o imperador chinês Shi Huangdi ordenou a destruição de todos os livros em 213 a.C. só para construir uma biblioteca que refletia os seus interesses. O imperador romano Carlos Magno construiu em 781, na sua corte em Aachen, uma biblioteca com grande parte de livros obtidos da Biblioteca Imperial, em Constantinopla. Muitos impérios europeus, incluindo o vaticano, foram erguidos no mesmo processo ambicioso (WRIGHT, 2014).

A democratização do acesso e o conhecimento compartilhado têm sido processos evolutivos. Embora os acervos individuais da Antiguidade visassem, de algum modo, o acesso ao conhecimento, constituíam-se numa realidade diferente. Tal realidade foi contrastada pela apologia social e científica das primeiras redes universais do saber da Idade Moderna, citadas por Barreto (2008), como a Academia de Lince de 1603; a *Royal Society* de 1660 (reconhecida em 1662); as Academias de Ciências de Londres, de Paris e de Berlim, criadas em 1665, 1666 e 1700, respectivamente, e alguns jornais científicos.

Para Barreto (2008), a consagração das redes do saber justifica-se na ideologia sobre o fluxo de informação e sua distribuição ampliada e equitativa, que está na origem da própria Ciência da Informação e dos ambientes híbridos de linguagem como a *Web*.

A *Royal Society*, por exemplo, desde a sua origem reuniu no mesmo espaço pesquisas de filósofos naturais como Isaac Newton e pesquisas modernas, tornando-se num ambiente de colaboração e comunicação científica aberto a todas as formas de conhecimento. Por isso, em 1663 foi reconhecida como a *Royal Society* de Londres para aprimorar o conhecimento. No concernente à distribuição ampliada do conhecimento, destaca-se o jornal *Philosophical Transactions* que estabeleceu os conceitos importantes de prioridade científica e revisão por pares, publicando o princípio matemático de Isaac Newton, a experiência de Benjamin Franklin sobre a natureza elétrica de um raio, sobre o motor de diferença de Charles Babbage, sobre o

desencadeamento do átomo de James Chadwick, entre outras. Atualmente a *Royal Society* aloca cerca de 42 milhões de euros por ano para encontros, eventos, parcerias, publicações, bolsas e prêmios, no âmbito da colaboração científica de cerca de 1600 cientistas espalhados pelo mundo (ROYAL SOCIETY, 2016).

Voltando à distribuição ampliada e equitativa, se olharmos para a semântica, com base na leitura de Hjørland (2007), como conjunto de esquemas, regras, sistemas e linguagens para atribuir significado à informação no âmbito da organização do conhecimento, subscrevemos a afirmação de Barreto (2008, p. 3) de que “o ideal do acesso ao conhecimento livre e para todos não surgiu com a *Internet*”. Aliás, este constitui o fulcro dos fundamentos que pretendemos demonstrar a partir de pensamentos anteriores à *Web*. Um dos pensadores sobre o acesso livre e/ou conhecimento compartilhado é Paul Otlet.

O internacionalista Belga Paul Otlet é considerado uma das figuras proeminentes na organização do conhecimento. Otlet desenvolveu vários esquemas para organizar a recolha e promover o acesso universal ao conhecimento humano através de uma rede global de informações, o *Mundaneum*<sup>2</sup>. O projeto de Otlet consistia na catalogação de recursos de informação (livros, revistas, jornais, fotografias, cartazes, panfletos, peças de museu e outra parte da propriedade intelectual), e no seu auge, segundo Wright (2014), registrou mais de 15 milhões de entradas de cartões de índice individuais.

As abordagens atuais da semântica sobre o processo de organização do conhecimento na *Web* podem ser vislumbradas a partir das concepções de Otlet sobre as descrições e análises de itens informacionais e a sua distribuição para o público mais amplo. É importante realçar que não estamos a afirmar que Otlet trabalhou sobre os aspectos da *Web Semântica*, tampouco que inventou a *Web*. Porém, a maioria das suas pesquisas encontra-se refletida na terminologia atual da *Web*. O fundamento para este argumento pode ser encontrado em Wright (2014, p. 8) ao considerar que:

em 1934 Otlet descreveu a sua visão sobre um sistema de computadores em rede chamado “telescópios elétricos” que permitiria que as pessoas procurassem através de milhões de documentos interligados, imagens e arquivos de áudio e vídeo. Para tal, cada pessoa estaria numa estação de trabalho equipada com uma interface de visualização e várias superfícies móveis ligadas a um repositório

---

<sup>2</sup> Atualmente é uma organização sem fins lucrativos, situada em Mons – Bélgica, que é dedicada à preservação e exposição do legado de Paul Otlet e Henri La Fontaine.

**Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v.3, n.1, p.138-158, jan./jun. 2016.

central que iria fornecer o acesso a uma ampla gama de recursos em qualquer tema de interesse. O sistema uniria indivíduos, instituições e governos e apresentaria máquinas de seleção capazes de identificar com precisão uma determinada passagem ou fato individual em um documento armazenado em microfilme que seriam recuperados através de uma ferramenta de indexação ou da recuperação mecânica.

Mesmo sem usar a terminologia atual, Otlet já trabalhava sobre questões semânticas ao criar o microfilme para o armazenamento sequencial de dados, com sistemas complexos de organização, para profundas transformações sociais, culturais e políticos, à semelhança do aumento das capacidades humanas na proposta da atual *Web Semântica*.

O projeto de Otlet ultrapassava fronteiras do acesso através da visão sobre o progresso social, realização científica, iluminação espiritual coletiva e liberdade, propiciados pelo livre fluxo de informações através de uma rede distribuída. Ao pensar sobre dados conectados e redes de usuários no âmbito do compartilhamento, vislumbrava-se o cenário informacional atual e complexo em que a figura do usuário final, que no passado se caracterizou pela aparente passividade nos ambientes de difusão, assume contornos conflitantes, na medida em que se encontra imiscuída em todas as etapas do ciclo informacional que, no entendimento de Le Coadic (2004), envolve a produção, a comunicação e o uso da informação. Atualmente existe uma dificuldade em identificar singularmente os usuários finais, já que são os mesmos produtores, mediadores, organizadores e agentes dos diferentes usos que se dão nos processos de apropriação. Por exemplo, no âmbito da produção do conhecimento, os usuários recorrem a diferentes fontes de informação para subsidiar as suas pesquisas e, ao mesmo tempo, atuam como mediadores no tratamento da informação, elaborando resumos, determinando os tópicos e fixando as palavras-chave. Consequentemente, influenciam os resultados da recuperação principalmente quando se usam os produtos de índice *Keyword in context (KWIC)* de Luhn (1960) na indexação automática.

A atual complexidade dos ambientes informacionais e da figura do usuário resulta das aspirações da própria humanidade sobre o universo do conhecimento, testemunhada pelo projeto da *Web Semântica* enquanto paradigma para o aumento das capacidades humanas em prol da informação.

Sendo Paul Otlet uma das figuras proeminentes na organização do conhecimento por meio de enfoques cujos alcances se observam na *Web*, porque o seu trabalho pouco aparece no contexto da sua criação? A ausência de Otlet na história da *Web* pode ser constatada na obra de

Castells (2003), ao iniciar a sua contextualização a partir do projeto de rede interativa de computadores da *Advanced Research Projects Agency (ARPA)* em 1969, inicialmente com fins militares. Entre vários fatores, o auge do projeto foi em 1989, com a proposição da *World Wide Web* pelo inglês Berners-Lee, a partir das propostas de Vannevar Bush sobre o *Memex*, da síntese do *on-line system* de Engelbart; do *Xanadu*, o hipertexto aberto e autoevolutivo de interligação de informações desenvolvido por Ted Nelson, bem como do sistema *hypercard* de interconexão de informações de Bill Atkinson (1987). O autor também destaca a influência da Microsoft, com a introdução do *software Windows 95* e do navegador *Internet Explorer* em 1995 e dos movimentos estudantis de produção de softwares abertos, de redes de computadores e de distribuição de protocolos de comunicação.

De salientar que, embora a publicação formal da *Web* tenha ocorrido em 1990, o seu desenvolvimento pode ser sintetizado a partir de 1980, quando Berners-Lee escreveu o programa *Enquire* que, contrariamente aos bancos de dados, armazenava associações aleatórias de informações (CASTELLS, 2003).

A mesma constatação acerca da ausência de Otlet na história da *Web* permanece na percepção de Berners-Lee (1996), ao destacar os princípios de design de software para a rede e os critérios para o desenvolvimento da *Web*.

A importância do projeto de Otlet não consistia apenas na coleção, mas nas visões e revolucionários processos de organização e acesso da informação. Talvez isso tenha sido percebido pelo estudante australiano Warden Boyd Rayward, ao dedicar maior parte das suas pesquisas sobre Otlet. Para Rayward (1994), foi Otlet que desenvolveu sistemas complexos de organização para integrar dados bibliográficos, imagens e textos que mais tarde impulsionaram e foram adaptadas para o hipertexto. Mesmo sem usar a terminologia hipertexto, Otlet pode ser considerado um dos precursores da *Web*, na medida em que em 1893 já levantava a necessidade de um sistema internacional de tratamento de informações que envolvia o uso de catálogos para novas formas de publicação. O sistema também incluía a gestão de bibliotecas, arquivos e museus como agências inter-relacionadas de desenvolvimento colaborativo, a criação de uma enciclopédia universal e da rede universal para aquilo que ele considerava como todo o conhecimento humano.

A fragmentação do hipertexto em nós para ser recuperado por vínculos associativos, que é

a tônica da *Web Semântica*, pode ser relacionada à Otlet, a partir das quatro categorias de informação (fatos, interpretação de fatos, estatísticas e fontes) com base nas quais cada artigo ou capítulo deveria ser analisado, para ser representado em cartões que formariam o catálogo. A partir da Classificação Decimal Universal (CDU) e com base no Instituto Internacional de Bibliografia, Otlet e Henri La Fontaine criaram o primeiro banco de dados de cartões que além de arquivos textuais, também classificaram imagens, estabelecendo a primeira enciclopédia que pretendia reunir todo o conhecimento humano, os padrões de publicação e o sistema complexo de indexação da informação (RAYWARD, 1994).

Para Marchand, Guérin e Barthès (1996), a história do hipertexto é atribuída aos projetos *Memex* de Bush (1945), aumento de Engelbart (1962) e *Xanadu* de Nelson (1960), mas o internacionalista belga expôs o seu trabalho em 1934, portanto, onze anos antes do artigo “*as we my think*” de Bush, de 1945. Judge (2001), por sua vez, considera que a *World Wide Web* incorporou a documentação que já estava organizada como hipertexto com base em metadados comuns por Otlet, e os vínculos associativos tornaram-se clicáveis pela conversão dos respectivos bancos de dados para o acesso à *Web* a partir de 1996.

Ao longo desta abordagem temos demonstrado que as questões que dizem respeito à *Web Semântica* têm uma estrita relação com a organização e o compartilhamento do conhecimento. Quando em 1962 Douglas Engelbart publicou o artigo “*Augmenting Human Intellect: a Conceptual Framework*”, mesmo enaltecendo os ganhos alcançados durante anos, demonstrava claramente as dimensões que o uso de computadores ainda não tinha alcançado. Através da inovação descontínua que quebra os paradigmas nos quais determinados padrões sociais estejam assentes, Engelbart (2003) propunha a computação interativa de modo a tornar colaborativa a construção do conhecimento e “aumentar” a capacidade humana na solução de problemas.

Engelbart (2003) justificava o seu fundamento a partir da natureza do processo de produção do conhecimento que não é individual, mas do grupo. Para tal, havia a necessidade de tornar a informação mais acessível e a interação amigável para proporcionar outras atividades humanas no uso de computadores. O *online system – NLS*, desenvolvido entre 1960 e 1970 no seu próprio laboratório “*Augmentation Research Center*”, é prova das suas aspirações. Este sistema facilitou a criação de bibliotecas digitais, o armazenamento e recuperação de documentos eletrônicos utilizando o hipertexto, a criação de novas interfaces gráficas que permitiram a leitura

de e-mails, novas opções de processamento de texto e vídeo teleconferência. Em 1969 o sistema *online* foi aplicado como um dos nós da *Advanced Research Projects Agency commissions (ARPANET)* e os seus conceitos continuam em uso atualmente na *Web*.

A leitura do artigo de Engelbart (2003) permite-nos concluir que, à semelhança da obra de Otlet, se trata de alguma das primeiras reflexões atinentes à integração de dados para dar maior significado à informação e realizar negócios na *Web*, tal como a *Web Semântica* se apresenta atualmente. Estas ideias se desdobram numa continuidade. O artigo do próprio Berners-Lee de 1996 é prova disto, ao apontar os três desafios da *Web* para o futuro: a melhoria da infraestrutura, o aumento da *Web* como meio de comunicação, e interação entre as pessoas e o uso de dados compreensíveis por máquinas para permitir a análise e resolução de problemas dos humanos. Por isso que a *Web Semântica*, segundo Berners-Lee, Hendler e Lassila (2001), é extensão da *Web* atual, com a proposta de maior integração de dados compreensíveis tanto para computadores como para humanos, de modo a contrapor se à quebra de informações, garantindo a interoperabilidade ou reutilização de dados e a simbiose humano – computador na solução de problemas.

Na Ciência da Informação, a integração de dados é um assunto que vem sendo desenvolvido desde os tradicionais esquemas de organização de conhecimento como coleções de dados bibliográficos. Este é o fundamento que pretendemos abordar a seguir.

## **2.2 A ORGANIZAÇÃO DO CONHECIMENTO**

Na Ciência da Informação, a organização do conhecimento envolve as abordagens tanto de caráter epistemológico, como prático sobre a produção, o tratamento, a preservação, a representação, a difusão e a recuperação da informação. A sua conceituação, de acordo com Hjørland (2008), abrange o sentido restrito e o sentido amplo. No sentido restrito, envolve as atividades relacionadas com a natureza e a qualidade dos processos e sistemas de organização do conhecimento, aplicados tanto na organização como na representação de documentos, obras e conceitos. No sentido amplo, a organização do conhecimento vai desde a divisão das áreas de conhecimento, instituições de pesquisa, disciplinas, profissões até o ciclo no qual cada setor ou área permeia a produção, difusão e uso do conhecimento.

Conforme Hjørland (2008) acrescenta, a abordagem restrita da organização do

conhecimento não pode subsistir sem o sentido amplo porque o universo do conhecimento constitui um todo. Quanto a nós, aqui se justifica o caráter interdisciplinar do campo e se fundamenta o olhar simbiótico entre a Ciência da Informação e outras áreas que trabalham sobre o mesmo objeto. Por exemplo, acreditamos que a Ciência da Informação deveria abordar a *Web Semântica* com base na premissa geral da organização do conhecimento, criando intersecções com outros campos, para responder aos diversos questionamentos levantados no seu próprio objeto, majoritariamente enleados à produção, difusão e uso da informação.

Se olharmos as bibliotecas como primeiras unidades ou ambientes informacionais cujo papel estava vocacionado à organização e distribuição equitativa da informação, perceberemos que a classificação se posiciona no rol das abordagens pioneiras sobre a organização do conhecimento no sentido restrito. Porém, conforme Barbosa (1969) observa, nem sempre as bibliotecas foram vistas como organismos vivos. As bibliotecas da antiguidade armazenavam os livros com o intuito de preservá-los e, conseqüentemente, eram pouco consultados. Barbosa (1969) acrescenta que isso só mudou no século XIX, com o surgimento das bibliotecas universitárias e públicas. Essa mudança situa-se na abordagem da organização do conhecimento naquilo que Guimarães (2015) denomina de processo helicoidal ininterrupto, ou seja, a necessidade de organizar o conhecimento para que este possa estar acessível e ser apropriado pela sociedade gerando novo conhecimento em um ciclo. Um dos pontos fulcrais da referida mudança, em termos da institucionalização, ocorreu no final do século XX, com a criação da *International Society for Knowledge Organization – ISKO* que passou a reunir mais de 600 pesquisadores empenhados em várias questões do campo da organização do conhecimento.

Os sistemas de classificação, segundo Barbosa (1969), remontam da Filosofia desde Platão (428-347 A.C.) que dividiu o conhecimento em física, ética e lógica, passando por Porfírio (305 D.C.) que usou a árvore de assuntos gerais para os específicos. James Duff Brown (1939) agrupou o conhecimento em quatro campos. Melvil Dewey (1876) baseou-se no sistema de Francis Bacon (1605) para enquadrar o conhecimento humano na memória, imaginação e razão formando, deste modo, as dez categorias da Classificação Decimal, voltadas para o gerenciamento de coleções de bibliotecas. A estes, Hjørland (2008) acrescenta a classificação de Henry Bliss (1940), baseada na ordem do conhecimento conforme as descobertas das ciências, na abordagem tradicional da organização do conhecimento.

A Classificação Decimal de Dewey foi adaptada por Otlet, na Classificação Decimal Universal, com o emprego de sinais auxiliares para representar assuntos complexos nos códigos ou produtos da análise documentária. A *Library of Congress*, por sua vez, adotou um sistema de classificação baseado nas suas próprias necessidades (BARBOSA, 1969).

Ao lado da classificação, a catalogação também vem desempenhando um papel preponderante na representação descritiva no âmbito do tratamento documentário. Algumas das suas vicissitudes foram influenciadas pela introdução dos computadores na recuperação entre 1957 e 1959 e pela emergência de catálogos públicos *online* em meados de 1980. Chu (2007) observa que, ao lado dos computadores pessoais, os catálogos públicos *online* foram importantes na desmistificação da complexidade da recuperação e no significado do usuário final. Antes, a recuperação da informação era feita em terminais de impressão, com interfaces não amigáveis, e com a ajuda de profissionais de informação, o que acarretava mais custos e demanda de tempo.

A organização do conhecimento no âmbito da representação descritiva também foi influenciada pelo surgimento de grupos voltados à discussão e implementação de metadados que permitem a descrição, a localização, a preservação, a interoperabilidade, a administração, a gestão de direitos autorais, etc., através de padrões de metadados como *Machine Readable Cataloging – MARC e Dublin Core – DC*. De igual modo, a *Anglo-American Cataloguing Rules (AACR2)* destaca-se pelas regras de descrição que garantem a unicidade na identificação, localização e recuperação de recursos em catálogos. Outras iniciativas estendem-se para o novo padrão *Resource Description and Access (RDA)* que pretende substituir as regras da AACR2 e modificar o formato *MARC* na descrição e acesso de recursos em diversos formatos de mídia. O *RDA* visa principalmente às bibliotecas, mas o seu uso é extensivo a outras comunidades (arquivos, editores, museus), incluindo a *Web Semântica*, e está alinhado com os modelos conceituais de dados bibliográficos e de autoridade desenvolvidos pela *International Federation of Library Associations and Institutions (IFLA)*.

No âmbito da organização do conhecimento, os padrões de metadados como o formato *MARC* enquadram-se no modelo da *Web Semântica*, pois consistem num conjunto de códigos e designações de conteúdos definidos para codificar registros capazes de serem interpretados por máquina, com a finalidade principal de possibilitar o intercâmbio de dados. O padrão *Dublin Core* também se tornou importante na descrição de recursos eletrônicos, na interoperabilidade e

na especificação da propriedade intelectual que vai facilitar a ação dos agentes que coletam as informações a partir de diferentes fontes no âmbito da *Web Semântica*.

Na organização do conhecimento, a *Web Semântica* está centrada em três elementos, citados por Marchand, Guérin e Barthès (1996), que estão na base do hipertexto desde Otlet, passando por Bush, Engelbart, Nelson até Berners-Lee: a ubiquidade, a onipotência e a onisciência. A ubiquidade está relacionada ao elevado volume de informações, ao qual Otlet contrapôs-se criando a bibliologia, a ciência da documentação. Para tal, Otlet propôs um conjunto de máquinas interligadas para as seguintes operações: criação de documentos de forma individualizada para relação com outros dados, interligação dos documentos e dados em qualquer lugar e recuperação automática. Assim, a ubiquidade tanto em Otlet como nos outros autores consiste na organização da informação em grande escala e no apoio para o conhecimento acumulado ao longo de séculos, para tornar fácil e rápido o conteúdo em manipulação (MARCHAND, GUÉRIN, BARTHÈS, 1996).

A onipotência é outro elemento do hipertexto que transcende o texto. Marchand, Guérin e Barthès (1996) apontam que foi inicialmente desenvolvida por Otlet a ideia de apresentar os documentos com inscrições adicionais de ferramentas de visualização e manipulações a fim de se obter novas combinações de fatos e relações entre ideias. Com isto, o usuário deixaria de ser passivo e estabeleceria ligações ativas, adicionando anotações e conexões pessoais aos dados. Estas ligações requerem algum tipo de conhecimentos informáticos por parte dos usuários na *Web*; por isso, as camadas de interoperabilidade em relação à sintaxe de descrição e de representação de metadados sobre recursos da *Web Semântica*, aliadas à camada de ontologias, vêm reforçar esta tendência de ligações entre dados para aumentar o universo do conhecimento.

A onisciência é o último elemento do hipertexto, na sua forma inicial, que Marchand, Guérin e Barthès (1996) também atribuem a Otlet, ou seja, as operações visando à ubiquidade e onipotência seriam realizadas por um cérebro mecânico coletivo e autêntico. O conceito do hipertexto sempre esteve atrelado à cooperação na criação de *links* personalizados e comentários dentro da construção social do conhecimento. Conforme os autores acrescentam “um hipertexto é adaptável e partilhável, [...] nunca é um produto final, mas continua a ser, para os seus usuários, uma área de expressão e de memória que está em constante evolução”. Esta percepção estabelece a diferença entre as primeiras aspirações do hipertexto e a fase inicial da sua implementação na

*Web*. No início, a arquitetura da *Web* era de cliente-servidor e não permitia a comunicação, na medida em que o autor só tinha o domínio sobre determinado número de documentos, em relação aos quais podia criar *links* para outros documentos. Isso mudou com a *Web* colaborativa e a Wikipédia é um exemplo da construção colaborativa do conhecimento, pelas transformações do hipertexto e *links* de cada usuário.

No contexto da organização da informação para o acesso compartilhado em prol do aumento das capacidades humanas sobre o conhecimento, também se deve destacar as contribuições de alguns pesquisadores pioneiros na recuperação da informação. Entre tantos, Chu (2007) destaca os seguintes: Mortimer Taube, Hans Peter Luhn, Calvin N. Mooers, Gerard Salton e Karen Spärck Jones. Nas contribuições de Taube (1952), o autor salienta a indexação coordenada pela combinação de unitermos. Estes termos de conceitos permitiram o uso de tesouros como linguagens documentárias para expandir a representação e recuperação por meio de diferentes pontos de acesso, melhorar a interação usuário – sistema, reduzir a ambiguidade, e padronizar a terminologia em determinados ambientes controlados.

Dentre as diversas contribuições de Luhn (1947), apontadas por Chu (2007), destaca-se a combinação de métodos estatísticos para a produção automática de índices, resumos e a indexação automática, bem como a disseminação seletiva da informação conforme os perfis dos usuários. A indexação automática tornou-se um elemento-chave na atual fase da explosão informacional, pela limitação da capacidade humana para a coleta, análise, tratamento e difusão dos recursos informacionais de acordo com os respectivos assuntos.

Calvin Mooers (1950) cunhou o termo *Information Retrieval (IR)* e desenvolveu aplicações para o tratamento de texto não estruturado. Salton (1968), por sua vez, desenvolveu o *System for the Manipulation And Retrieval of Texts (SMART)*, baseado no modelo vetorial para a recuperação do texto completo. Já Jones (1972) desenvolveu métodos de processamento em linguagem natural através de dois métodos para a recuperação na *Web*, o *Term Frequency (TF)* e *Inverse Document Frequency (IDF)* (CHU, 2007).

Como observamos, as contribuições da recuperação da informação no âmbito da organização do conhecimento são imensas. Porém, na sua maioria, foram concentradas em abordagens práticas. Esta é uma das inquietudes levantadas por Hjørland (2008) sobre a ausência de pressupostos teóricos para confrontar diferentes abordagens relativas à organização do

conhecimento. Ademais, estas questões voltam à tona quando se trata de determinar, por meio de modelos teóricos, o sentido e os desdobramentos da semântica tratada na *Web* em relação ao campo da Ciência da Informação.

### 2.3 A SEMÂNTICA DA LINGUAGEM E DA WEB

O termo “semântica” assume diferentes sentidos em função do contexto em que é empregue. Geralmente, é associado a campos interdisciplinares da Filosofia, Linguística, Psicologia, Sociologia, Ciência da Computação e Ciência da Informação, como o estudo do significado da linguagem, tanto na relação entre os sujeitos da comunicação como na relação destes com os objetos.

Enquanto a sintaxe se preocupa com a estrutura ou com os padrões de relacionamento entre palavras ou itens, a semântica encarrega-se do seu significado através de conceitos ou da redução de ambiguidades nos processos de interpretação e apropriação consubstanciados pelas relações<sup>3</sup>. Além das palavras, a preocupação com o significado é extensiva às imagens e outros signos. Por isso, mesmo não tendo sido tratado neste trabalho, a semântica também está relacionada à semiótica.

Na Ciência da Informação, segundo Hjørland (2007), a semântica tem uma estrita relação com a organização do conhecimento no uso dos cabeçalhos de assunto, da classificação, de taxonomias, de esquemas de categorização, dos tesauros, das redes semânticas, e das ontologias. Contrariamente a Hodge (2000, apud HJØRLAND, 2007), o autor também inclui os sistemas de organização do conhecimento como arquivos de autoridade, mapas bibliométricos e bases de dados bibliográficos. Assim, a semântica na Ciência da Informação visa essencialmente à recuperação da informação e, por isso, tem o olhar interdisciplinar do processamento em linguagem natural.

Na linguagem, a semântica está relacionada ao léxico, à nomenclatura, ao vocabulário e à terminologia. O léxico é o conjunto de regras e relacionamentos semânticos entre palavras de uma determinada língua a partir dos respectivos conceitos. No caso da língua portuguesa, a rede

---

<sup>3</sup> Ao falarmos da redução de ambiguidades, partimos do pressuposto da complexidade da comunicação em que quase se torna impossível eliminar toda a ambiguidade possível, mesmo trabalhando com relações lógicas bem definidas.

léxica conceitual *WordNet.Pt*, segundo Marrafa et al. (2005), foi desenvolvida pelo Centro de Linguística da Universidade de Lisboa, com base do modelo da *EuroWordNet* para as línguas europeias e da *WordNet* da Princeton para o inglês. Trata-se de uma ontologia lexical da língua portuguesa que pode ser aplicada em várias áreas, desde a comunicação até a produção de termos e dos respectivos conceitos para a construção de linguagens artificiais em ambientes controlados.

Cintra et al. (2002) definem a nomenclatura como o processo de identificar algo pelo respectivo nome. Saussure (2006) afirma que, para certas pessoas, este é o princípio essencial da língua, e representa uma visão simplista porque o signo linguístico não une uma coisa a uma palavra. Pelo contrário, une um conceito, um significante e um significado. Aqui está exposta a ambiguidade ou arbitrariedade que o signo pode assumir na relação com o significado. Voltaremos a abordar estes aspectos mais adiante.

O vocabulário consiste na forma da linguagem usada em determinados contextos de comunicação ou por grupos específicos. Por exemplo, pode se falar do vocabulário de especialistas na saúde, no direito, na computação ou em artigos científicos. Daí que o vocabulário é uma linguagem de especialidade. A terminologia também se manifesta em ambientes ou grupos específicos.

O léxico, a nomenclatura, o vocabulário e a terminologia afetam os modos de apropriação da informação pela semântica. Conforme referimos anteriormente, a semântica é um assunto amplamente discutido na Linguística, principalmente quando se trata da língua como objeto. Como Saussure (2006) aponta, as línguas são dadas pelos respectivos grupos linguísticos e a principal forma de conhecê-las é por intermédio da escrita, ou seja, a escrita é a representação da língua. Equivale por outras palavras dizer que, na comunicação, o objeto linguístico dá-se pela fala e representa-se pela escrita. Aqui surge um dos problemas da comunicação relacionados com a semântica. Até que ponto o sistema de signos da escrita reflete o objeto linguístico construído no interior do ser cognoscente? A resposta para esta pergunta está pautada na teoria complexa da comunicação e não se pode aferir de forma taxativa.

Se a escrita propicia alguns dos problemas semânticos na exteriorização da informação, a situação se complica na recuperação da informação, em que o usuário se apropria de um simples termo de busca para exteriorizar a totalidade da sua necessidade informacional. Ademais, Wilson (1981) esclarece que existem usuários que não reconhecem as respectivas necessidades e, por

isso, se tornam incapazes de exteriorizá-las. Isso acontece porque o comportamento de busca não é consequência direta de uma necessidade, isto é, existem vários fatores que afetam a procura e que envolvem as competências informacionais através do grau da especificidade das estratégias de cada usuário.

O problema semântico está patente na maioria dos sistemas de representação. Uma das principais mudanças na representação ocorreu em 1848 com o colapso da ideologia iluminista que, segundo Harvey (2004), considerava a existência de um único modo correto de representação para controlar e organizar o mundo. Esta mudança trouxe novos contornos na representação. Por exemplo, através da teoria estruturalista da linguagem de Saussure, o sentido das palavras é determinado antes pela sua relação com outras palavras do que pela sua referência a objetos. Neste caso, a semântica resulta dessa relação e do contexto em que se estabelece a comunicação e não apenas pela relação significante – significado. Esse contexto é fornecido pelo tempo, pelo espaço, pela cultura, e por outras variáveis que condicionam determinados ambientes e circunstâncias de apropriação do conhecimento.

Na Ciência da Informação, os tesouros e outras linguagens documentárias refletem as especificidades da teoria estruturalista da linguagem com base no princípio da contextualização, a partir do qual se procura estabelecer a garantia literária. Conforme Gomes (1990), para diminuir a ambiguidade, considera-se o contexto geral da área no qual o termo está inserido e o contexto específico no qual o termo aparece. Assim, a garantia literária é determinada pelo contexto específico, ou seja, a semântica é dada pela terminologia adotada na literatura da área ou pelos usuários que o tesouro procura satisfazer. Por outras palavras, o significado dos descritores não é determinado pela relação ao objeto na linguagem natural, mas pela terminologia do domínio de conhecimento.

Até aqui demonstramos que a semântica está relacionada à linguagem. Por isso, ela se estende para todos os ambientes em que se usa a linguagem. A *Web* é um desses ambientes e, neste caso, os problemas semânticos tornam-se ainda complexos pela hibridização de linguagens. Além da linguagem escrita, o hipertexto foi concebido de modo a congrega no mesmo espaço a informação contínua de outras linguagens como sonora, simbólica, oral, entre outras. Daí que o entendimento da semântica na *Web* passa necessariamente pela percepção das estruturas destas linguagens e pela organização dos recursos informacionais.

A linguagem *Resource Description Framework (RDF)*, por exemplo, foi desenvolvida no âmbito da *Web Semântica* para criar modelos simples de dados usando a sintaxe para marcações *eXtensible Markup Language (XML)*, de modo a obter uma semântica formal independentemente da linguagem usada na comunicação. O mesmo se pode dizer em relação à *Web Ontology Language (OWL)* no âmbito do processamento do conteúdo semântico dos recursos a partir de ontologias.

De acordo com o que foi exposto, percebe-se que a *Web Semântica* está situada no contexto da organização do conhecimento, no papel da disponibilização, acesso, uso e reuso de dados, explorando as potencialidades dos computadores. A pergunta que se coloca é se a *Web Semântica* está inserida na linguagem de comunicação ou na sintática ou organização de dados? Warner (2007, apud HJØRLAND, 2008) observa que a tarefa da sintática pode ser transferida para máquinas, mas a semântica requer o envolvimento humano direto, pois está ligada às transformações do emprego dos signos em determinados ambientes ou contextos de comunicação. Portanto, percebemos que a semântica é um problema linguístico, cuja solução transcende a sintaxe ou a camada de relacionamento de dados.

Nós entendemos que a semântica na linguagem está relacionada aos problemas linguísticos de cognoscência humana, sendo por isso de difícil solução com base em inferências das máquinas. Já a organização pode nos levar a experiências interessantes como temos verificado com as redes neurais na Inteligência Artificial, com as associações e classificações dos algoritmos na mineração de dados ou até com as buscas associativas em lojas de *e-commerce* (FERNEDA, 2006; TAN, STEINBACH, KUMAR, 2009).

A questão formulada anteriormente remete-nos aos alcances que a Ciência da Informação poderá atingir em relação ao seu objeto, a partir de um olhar próprio lançado sobre a proposta da *Web Semântica*. O problema, segundo Hjørland (2007), está na ausência de referenciais teóricos consolidados que nos permitam fazer abordagens sobre a semântica na área. Por exemplo, dentro do campo existe a semântica na classificação que nos remete ao princípio da garantia literária, na recuperação da informação que está ligada aos métodos estatísticos de ocorrência de signos nos documentos, na bibliometria sobre relações de citação, etc. As abordagens sobre a semântica na Ciência da Informação estão ligadas ao uso e não nos permitem visualizar num único modelo as dimensões da *Web Semântica*, pois estão fragmentadas.

Os problemas que enfrentamos na Ciência da Informação não nos devem alarmar, conforme nos tranquiliza Wersig (1991), ao considerar o estudo do conhecimento pós-moderno como o objeto do campo. Segundo o autor, ao contrário das ciências clássicas cuja gênese está enraizada na busca do entendimento completo de como o mundo funciona, a Ciência da Informação nasce na necessidade de desenvolver estratégias para resolver os problemas causados por ciências e tecnologias do clássico. Por isso, na maioria das vezes, está destituída de referenciais teóricos para fundamentar as suas abordagens.

Apesar dos questionamentos colocados, dúvidas não se levantam em relação ao papel da *Web Semântica* na organização do conhecimento. A nossa observação se baseia no pressuposto de que a sintaxe emana da semântica, pois qualquer esquema de organização depende da estrutura cognitiva que permite a sua criação e compreensão. Por exemplo, é a semântica que nos permite organizar as frases, localizar o predicado e identificar a respectiva função numa frase. O mesmo raciocínio se aplica na *Web Semântica* independentemente do tipo de semântica nela tratada e dos respectivos desdobramentos para a Ciência da Informação.

Acreditamos que a semântica da linguagem traduz a essência da natureza complexa da comunicação humana. Ela não é fruto das tecnologias de informação e comunicação, pois é anterior a elas<sup>4</sup>. Porém, através delas ganha maiores proporções no ciclo informacional e no volume de informações. Não se trata de problema dos meios, mas dos fins. É um problema herdado das estruturas, modelos e contextos complexos que permeiam a nossa comunicação.

### 3 CONSIDERAÇÕES FINAIS

Conforme nos propusemos a referenciar, a organização do conhecimento é uma área interdisciplinar. Na Ciência da Informação, foi desenvolvida num enfoque prático herdado da Biblioteconomia e Documentação, na maioria das vezes, voltado para a recuperação da informação. A sua gênese é anterior ao período pós-guerra da institucionalização da Ciência da Informação como campo científico. O seu desenvolvimento foi fortemente influenciado por Paul Otlet que, mesmo sem usar a maioria das terminologias atuais, contribuiu para o atual estado de arte.

---

<sup>4</sup> É o que Levy (2007) observa, ao definir o virtual como conceito abstrato de significação da linguagem. Enquanto os computadores manipulam os signos, a semântica permanece na mente humana.

*Informação & Tecnologia (ITEC)*, Marília/João Pessoa, v.3, n.1, p.138-158, jan./jun. 2016.

A relação entre a organização do conhecimento e a semântica se constitui a partir do “processo helicoidal ininterrupto” nela subjacente, aliado à complexidade tanto dos ambientes de produção, tratamento e difusão como do processo de transferência da informação.

Em relação à semântica prevalecente na *Web*, o desenvolvimento da Ciência da Informação sob o olhar prático da solução de problemas contribui para o modelo pragmático e fragmentado pelo qual observamos a maioria dos seus fundamentos, especificamente em relação à recuperação da informação. Por isso, cremos que é necessário fortalecermos a interdisciplinaridade, construindo modelos teóricos que nos permitam analisar a *Web Semântica* em diferentes perspectivas da informação. Entre outros aspectos, estes modelos vão nos permitir visualizar a semântica na *Web* no contexto da Ciência da informação, criando intersecções com outros campos e a partir das especificidades de cada linguagem que compõe o hipertexto.

### ***From Paul Otlet to the Semantic Web: theoretical contributions on the organization of knowledge***

#### **ABSTRACT**

This work consists of a qualitative approach, basic in nature and analyzes the interdisciplinary field of knowledge organization based on the literature. In this context, we revisit some theoretical contributions from the bibliographic classification systems, through the Paul Otlet Treaty of Documentation and the interactive computing Engelbart to the Semantic Web. While mapping a look about the semantics in production environments, diffusion and information appropriation, also questions the models highlighted in Information Science that form the basis for understanding the inherent semantic issues in information as an object of field of study. It concludes that the Information Science developed from the practical, and the lack of theoretical models consolidated in the area hampers the perception of the Semantic Web reaches for the object “information” from the interdisciplinary context. Semantics is linked to language, the complexity of human communication and precedes syntax. Therefore, regardless of the semantic Web treated in kind, their contribution is the knowledge organization, for any organization scheme depends on the cognitive structure that enables the creation and understanding. Undoubtedly, this approach will contribute to the discussions on the knowledge organization within the Information Science, rescuing Otlet figure that is sometimes ignored in the literature. Similarly, it will subsidize the foundations of the Semantic Web from a specific look of Information Science.

**Keywords:** Information Science. Semantics. Semantic Web. Knowledge organization. Informational sharing.

## REFERÊNCIAS

BARBOSA, A. P. **Teoria e prática dos sistemas de classificação bibliográfica**. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969.

BARRETO, A. A. Uma quase história da ciência da informação. **Datagramazero - Revista de Ciência da Informação**, v. 9, n. 2, 2008.

Disponível em: <[http://www.dgz.org.br/abr08/Art\\_01.htm](http://www.dgz.org.br/abr08/Art_01.htm)>. Acesso em: 17 abr. 2013

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: A new of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, p. 35-43, 2001.

BERNERS-LEE, T. The World Wide Web: past, present and future. **Computer Magazine**, v. 29, n. 10, 1996. Disponível em: <<https://www.computer.org/csdl/mags/co/1996/10/rx069-abs.html>>. Acesso em: 2 set. 2014.

CHU, H. **Information representation and retrieval in the digital age**. 3ª Tiragem, New Jersey: Asist&T, 2007.

CINTRA, A. M. M. et al. **Para entender as linguagens documentárias**. 2.ed. São Paulo: Polis, 2002.

ENGELBART, D. C. Improving our ability to improve: a call for investment in a new future. **Simpósio Co-evolution da IBM**, Set. 2003. Disponível em:

<[http://www.almaden.ibm.com/coevolution/pdf/engelbart\\_paper.pdf](http://www.almaden.ibm.com/coevolution/pdf/engelbart_paper.pdf)>. Acesso em: 04 mai. 2013.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação.

**Ciência da Informação**, v. 35, n. 1, p. 25-30, 2006. Disponível em:

<<http://www.scielo.br/pdf/ci/v35n1/v35n1a03.pdf>>. Acesso em: 27 jun. 2016.

FONTOURA, M. C. **A documentação de Paul Otlet: uma proposta para a organização racional da produção intelectual do homem**. 220f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília-DF, 2012.

GOMES, H. E. **Manual de elaboração de tesouros monolíngues**. Brasília: Programa Nacional de Bibliotecas de Instituições de Ensino Superior, 1990.

GUIMARÃES, J. A. C. Organização do conhecimento: passado, presente e futuro em um contexto de diversidade cultural. In: GUIMARÃES, J. A. C e DODEBEI, V. (Orgs.)

**Organização do conhecimento e diversidade cultural**. Marília: ISKO Brasil, Fundepe, 2015. p.13-19.

HARVEY, D. **Condição pós-moderna: uma pesquisa sobre as origens da mudança cultural**.

Tradução de SOBRAL, A. U; GONÇALVES, M. S.: São Paulo: Edições Loyola, 2004.

HJØRLAND, B. Semantics and knowledge organization. **Annual Review of Information Science and Technology**, p.367-405, 2007.

HJØRLAND, B. What is Knowledge Organization (KO)? **Knowledge Organization**, v.35, n.2/3, p. 86-101, 2008. Disponível em: <[https://is.muni.cz/el/1421/podzim2008/VIKMA13/um/6341865/Hjorland - What is Knowledge Organization.pdf](https://is.muni.cz/el/1421/podzim2008/VIKMA13/um/6341865/Hjorland_-_What_is_Knowledge_Organization.pdf)>. Acesso em: 30 jun. 2016.

JUDGE, A. **Union of International Associations - Virtual Organization: Paul Otlet's 100-year Hypertext Conundrum?** Disponível em: <<https://www.laetusinpraesens.org/docs/otlethyp.php>>. Acesso em 18 jul. 2016.

LE COADIC, Y. F. **A Ciência da Informação**. Brasília, DF: Briquet de Lemos/Livros, 2004.

LÉVY, P. **O que é virtual: fronteiras do pensamento**. Brasil, 2007. Documentário Arroba. Direção de Hique Montanari.

LUHN, H. P. Keyword-in-context index for technical literature. **American Documentation**, v.11, n.4, p. 288-295, 1960.

MARCHAND, Y., GUÉRIN, J. L., BARTHÈS, J. P. A. From a set of technical documents to a hypertext system on the Web. **World Conference of the Web Society - WebNet**, AACE, 1996.

RAYWARD, W. B. Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext. **Journal of the American Society for Information Science**, v. 45, n.4, 1994, p.235-250.

ROYAL SOCIETY. Disponível em: <<https://royalsociety.org/about-us/history/?from=basefeature#timeline>>. Acesso em: 8 jul. 2016.

SANTAREM SEGUNDO, J. E. A documentação: Paul Otlet e o facebook. In: CRIPPA, G.; MOSTAFA, S. P. **Ciência da informação e documentação**. Campinas: Editora Alínea, p. 99-112, 2011.

SAUSSURE, F. de. **Curso de linguística geral**. 27.ed. São Paulo: Cultrix, 2006.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining - Mineração de dados**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009

WILSON, T. D. On user studies and information needs. **Journal of Documentation**, v.37, n.1, p.3-15, 1981. Disponível em: <<http://www.informationr.net/tdw/publ/papers/1981infoneeds.html>>. Acesso em: 2 mar. 2011.

WRIGHT, A. **Cataloging the world: Paul Otlet and the birth of the information age**. Nova Iorque: Oxford University Press, 2014.