

TRADUÇÃO PARA PORTUGUÊS BRASILEIRO

HJØRLAND, Birger. Classification. *Knowledge Organization*, Würzburg, v. 30, n. 2, p. 87-111, 2003.

Birger Hjørland

Escola Real de Biblioteconomia e Ciência da Informação, Copenhagen Email: bh@db.dk

Birger Hjørland é professor de Ciência da Informação (CI) na Royal School of Library and Information Science em Copenhagen, Dinamarca. Ele foi professor de CI na College University of Borås, Suécia. Ele tem um Ph.D. em Ciência da Informação pela Universidade de Gothenburg, Suécia e mestrado em artes em psicologia pela Universidade de Copenhagen, Dinamarca. Por doze anos, o professor Hjørland trabalhou como bibliotecário pesquisador na Royal Library em Copenhagen, onde era um especialista em psicologia e coordenador dos serviços de referência baseados em computador da biblioteca. Ele ensinou História e Filosofia da psicologia na Universidade de Copenhagen e também ensinou Recuperação de Informação no Instituto de Linguística Aplicada e Matemática em Copenhagen. Apesar da experiência do Professor Hjørland no campo da psicologia, seu trabalho é muito influenciado e orientado por perspectivas filosóficas e sociológicas. Ele é membro da ASIS & T e ISKO e é o autor de *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science*.

RESUMO

Este artigo está organizado em 10 seções: (1) Organização do Conhecimento (OC) é um amplo campo interdisciplinar, muito mais amplo do que Library and Information Science (LIS). (2) Dentro da LIS, há muitas abordagens e tradições diferentes de OC com pouca influência mútua. Essas tradições foram, em grande medida, definidas por novas tecnologias, razão pela qual a integração teórica e a sustentação não foram bem consideradas. As tradições mais importantes voltadas para a tecnologia são: a) Indexação e classificação manual em bibliotecas e obras de referência, b) Documentação e comunicação científica, c) Armazenamento e recuperação de informação por computador, d) OC baseado em citações e e) Texto completo, hipertexto e abordagens baseadas na Internet. Essas tradições, tomadas em conjunto, definem muito o foco especial da LIS na OC. Para OC como um campo de pesquisa, é importante estabelecer um quadro teórico frutífero de referência para esse grande campo. Este artigo fornece algumas sugestões. (3) Uma distinção teórica importante a considerar é aquela entre as formas sociais e intelectuais de OC. As formas sociais de OC estão relacionadas à formação profissional, disciplinas e grupos sociais, enquanto a

¹ Tradução e publicação autorizadas pelo autor.

² Mestranda em Ciência da Informação pela Universidade Federal de Pernambuco, Graduada em Biblioteconomia pela Universidade Federal de Pernambuco (2019), atuando principalmente nos seguintes temas: Organização da Informação e Filosofia da Linguagem de Mikhail Bakhtin (gêneros do discurso e indexação). <https://orcid.org/0000-0003-2213-5742>

organização intelectual está relacionada a conceitos e teorias nos campos a serem organizados. (4) A perspectiva social inclui, tanto os sistemas de gêneros e documentos, bem como o sistema social de produtores de conhecimento, intermediários e usuários de conhecimento. (5) Este sistema social de documentos, gêneros e agentes disponibiliza uma estrutura muito complicada de potenciais pontos de acesso do assunto (subject access points - SAPs), que podem ser usados na recuperação de informação (RI). O objetivo básico da pesquisa em OC é desenvolver conhecimentos sobre como otimizar este sistema de SAPs e sua utilização em RI. (6) SAPs podem ser vistos como signos, e sua produção e uso podem ser entendidos de um ponto de vista semiótico social. (7) O conceito de paradigmas também é útil porque diferentes grupos e interesses tendem a ser organizados de acordo com um paradigma e a desenvolver diferentes critérios de relevância e, portanto, diferentes critérios de semelhança em OC. (8) A unidade básica em OC é a relação semântica entre dois conceitos, e tais relações estão embutidas em teorias. 9) Na classificação, coisas semelhantes são agrupadas, mas o que é considerado semelhante não é uma questão trivial. (10) O artigo conclui com a consideração de métodos para OC. Basicamente, os métodos de qualquer campo estão ligados a teorias epistemológicas. Este também é o caso com OC. Os métodos existentes, conforme descritos na literatura de OC, se encaixam em uma classificação de visões epistemológicas básicas. O debate sobre os métodos de OC no nível mais profundo, portanto, implica uma discussão epistemológica.

1 O que é organização do conhecimento (OC)?

Na comunidade de *Library and Information Science* (LIS), Organização do Conhecimento (OC) significa, especialmente, a organização da informação em registros bibliográficos, incluindo índices de citações, registros de texto completo e a Internet. Ciência da Informação (CI) é basicamente sobre a melhor maneira de construir tais registros bibliográficos (o que é feito em OC), bem como a maneira ideal de utilizar determinados registros (o que é feito em RI de recuperação de informação).

A OC é, no entanto, um conceito muito mais amplo. O conhecimento é organizado em, entre outras coisas:

- A divisão social do trabalho (por exemplo, em disciplinas)
- Instituições sociais (por exemplo, em universidades)
- Linguagens e sistemas simbólicos
- Sistemas e teorias conceituais
- Literaturas e gêneros

Library and Information Science (LIS) ou apenas Ciência da Informação (CI), frequentemente ignorou este significado mais amplo de OC e, portanto, falhou em se basear em tais teorias mais amplas; ou quaisquer teorias. Minha visão central é que essas visões mais amplas não podem ser ignoradas. Qualquer tentativa de desenvolver princípios frutíferos para OC em LIS deve ser baseada em teorias mais amplas de OC.

Os principais atores em CI são os produtores de conhecimento (por exemplo, autores), os usuários e os intermediários. É sua interação com registros bibliográficos que é o foco da CI (LIS). Cada um desses atores traz consigo certos pré-entendimentos, visões, conceitos e linguagens adquiridas principalmente durante a socialização em sociedade. O sucesso da interação depende desses pré-entendimentos, conceitos e linguagens. Isso se aplica não apenas à "correspondência" de conceitos, mas especialmente à sua capacidade

de apoiar as tarefas dos usuários. Uma teoria dessas formas mais amplas de OC é, portanto, essencial para construir sistemas eficientes para OC em LIS.

A LIS reivindicou o status de campo de estudo por volta de 1876, quando Melvin Dewey publicou pela primeira vez seu sistema de classificação e quando escolas de biblioteconomia começaram a surgir. Quais são os resultados acumulados de mais de um século de pesquisa em OC? Entre os resultados estão uma série de normas e diretrizes, alguns desenvolvimentos teóricos, por exemplo, as regras de Cutter (1904), avanços na *análise de facetas* e, claro, mudanças importantes na tecnologia da informação (TI). É, no entanto, difícil esboçar o progresso mais teórico e científico neste campo porque ele parece amplamente ateuórico e fragmentado e porque muitas linhas de pensamento diferentes parecem coexistir. Na minha opinião, OC carece de teorias sobre seus conceitos mais fundamentais, como:

- Conceitos
- Critérios para inclusão de classe
- Significado
- Indexação
- Relações semânticas
- Assuntos
- Pontos de acesso do assunto (“SAPs”) e assim por diante.

Também acho que OC carece de investigações sérias sobre os métodos de OC e a base metodológica de OC. Os desenvolvimentos na prática parecem muito mais influenciados pelo progresso em TI do que pelo progresso em OC propriamente dito. Às vezes, as pessoas parecem até esperar que a TI em si possa substituir as preocupações conceituais e teóricas. Em minha opinião, podemos ter a melhor visão geral do progresso em OC se considerarmos cinco estágios baseados na tecnologia no desenvolvimento de OC. Se isso for verdade, é a expressão de uma crise no campo: que o campo não é movido por desenvolvimentos em suas próprias pesquisas. Falta conhecimento acumulado, que pode ser transformado de uma plataforma de TI para outra.

2 Fases baseadas na tecnologia no desenvolvimento da Organização do Conhecimento (OC)

Em library and information science (LIS), o conceito de OC está conectado ao desenvolvimento de sistemas de classificação e indexação em bibliotecas, bibliografias e bancos de dados eletrônicos. Hjørland, (2000; Hjørland & Kyllèsbech Nielsen, 2001) descreveu cinco estágios impulsionados pela tecnologia no desenvolvimento de OC. Cada estágio não é substituído por estágios subsequentes, mas continua a influenciar o campo. Devido a esse fato, a tecnologia além desses estágios também pode ser vista como uma entre outras causas para o desenvolvimento de abordagens mais ou menos complementares ou concorrentes em OC.

As etapas são:

- a. *Indexação e classificação manual em bibliotecas e obras de referência*

A catalogação em bibliotecas e os princípios organizacionais em obras de referência remontam ao passado. Pode-se traçar os primórdios do uso da ordem alfabética até a antiguidade, embora ela não tenha sido geralmente adotada até a Idade Média e não fosse o único método de classificação possível (Daly, 1967). Porém, na Idade Média e no início dos tempos modernos, os catálogos provavelmente eram vistos como inventários projetados para controlar a coleção e para obter informações necessárias para novas aquisições, ao invés de ferramentas para pesquisas por assunto.

A profissionalização da classificação e indexação de livros em bibliotecas apareceu por volta de 1876 com a publicação da *Classificação Decimal de Dewey* e com a fundação de library schools (e, portanto, de “biblioteconomia”), e assim por diante. Charles A. Cutter (1837-1903), Melvin Dewey (1851-1931), Henry E. Bliss (1870-1955) e S. R. Ranganathan (1892-1972), entre outros, foram reconhecidos como fundadores.

É geralmente reconhecido que este estágio produziu princípios para OC que ainda são válidos e importantes. Entre eles estão as “regras de Cutter” e a abordagem de Ranganathan para a classificação dos campos de assuntos. É claro que outros princípios se tornaram obsoletos e alguns são controversos. Frohmann (1994), por exemplo, considerou a abordagem de Dewey muito prejudicial porque ignorou a natureza semiótica da classificação e apenas propôs um formalismo vazio: “Os assuntos de Dewey eram elementos de um sistema semiológico de software administrativo tecnoburocrático padronizado para a biblioteca em sua encarnação corporativa, ao invés de alta cultura” (p. 112-113). Deve-se acrescentar, entretanto, que há muito pouca discussão geral sobre a importância desses (e de outros) princípios. Eles foram considerados padrões ou cânones, em vez de princípios baseados em pesquisas, sujeitos a investigação, modificação e revisão.

b. Documentação e Comunicação científica

O movimento de documentação desenvolveu-se a partir de 1895. Os fundadores foram Paul Otlet (1868-1944) e Henri Lafontaine (1854-1943), que fundaram o *The International Institute of Bibliography*² e em 1905-1907 publicaram a primeira edição da *Classificação Decimal Universal*, CDU, que foi uma extensão da Classificação Decimal de Dewey projetada para uma análise mais profunda do assunto. O *American Documentation Institute* foi fundado em 1937 e, em 1968, mudou seu nome para *American Society for Information Science* (ASIS). É difícil descrever as diferenças exatas entre biblioteconomia e documentação. Documentalistas estavam menos interessados em bibliotecas e suas coleções; estavam mais interessados no controle bibliográfico, na comunicação científica e na documentação científica, nos serviços de informação à indústria e na utilização do conhecimento nos documentos. Implicitamente, eles compartilhavam uma visão de OC baseada no conhecimento do assunto e na literatura do assunto. Os documentalistas costumam se considerar mais voltados para o serviço, mais orientados para a tecnologia e mais avançados do que os bibliotecários. Enquanto os bibliotecários tradicionais costumavam ter uma orientação para as humanidades, os documentalistas eram principalmente afiliados à ciência, tecnologia e negócios. Eles indexaram artigos individuais em revistas e livros e desempenharam um papel central no estabelecimento de revistas internacionais de resumo³. Eles estavam menos interessados em manter livros para seu próprio bem ou para propósitos culturais amplos, e estavam altamente interessados em estabelecer serviços que pudessem estimular a aplicação do conhecimento para fins específicos. A base dos estudos do usuário (Bernal, 1948) e da bibliometria (por exemplo, Bradford, 1948) também faz parte deste estágio / tradição, que pode, antes de mais nada, ser caracterizado por uma abordagem de assunto mais específica, um nível mais profundo de indexação, mais ênfase na tecnologia

moderna e uma atitude mais científica em relação aos objetivos e problemas. Este estágio lançou as bases para o desenvolvimento posterior de bancos de dados bibliográficos online.

c. *Armazenamento e recuperação de informações por computadores (principalmente 1950)*

A informática⁴ tem um desenvolvimento próprio, que não deve ser confundido com o desenvolvimento da biblioteca, da documentação e da ciência da informação. Como disciplina teórica, a ciência da computação foi fundada na década de 1930, bem antes do surgimento do computador moderno. Os pais fundadores foram os lógicos Alonso Church (1903-1995), Kurt Gödel (1906-1978), Stephen C. Kleene (1909-1994), Emil Post (1897-1954) e Alan Turing (1912-1954). Esse trabalho inicial teve uma profunda influência tanto no desenvolvimento teórico quanto no aplicado da ciência da computação, mas essas pessoas não estão diretamente ligadas à história da LIS ou da OC, a meu ver. A ciência da computação tem diferentes programas educacionais, revistas científicas e conferências em comparação com a LIS e também tem objetivos um tanto diferentes em comparação com a LIS. Arbib, Kfoury e Moll (1981) escrevem que “*A ciência da computação busca fornecer uma base científica para o estudo do processamento da informação, a solução de problemas por algoritmos e o projeto e programação de computadores*”. Na minha opinião, o objetivo da LIS é otimizar a utilização do conhecimento documentado. O objetivo principal das bibliotecas é fornecer acesso físico e intelectual às informações. O acesso intelectual é fornecido pela organização dos próprios documentos físicos ou pela organização das representações dos documentos em catálogos, bibliografias e bancos de dados. A LIS não se concentra principalmente na construção de algoritmos, mas em informar as pessoas sobre documentos. A LIS deve ser mais aberta a diferentes visões, mais reflexiva e meta-orientada e demonstrar lacunas e incertezas no conhecimento para os usuários. Isso é muito diferente de fazer, por exemplo, um sistema especialista que tem um desempenho ótimo refletindo alguns modelos cognitivos generalizados.

Com o advento dos computadores na década de 1950, a LIS e a OC foram influenciadas por essa nova tecnologia, e muitas pessoas sentiram que o futuro da LIS deveria ser parte da ciência da computação. Há uma indicação de que os “cientistas da informação” não se consideravam parte da biblioteconomia ou lidando com documentos e suas representações (cf., Hjørland, 2000a). Eles se sentiam muito mais relacionados à ciência da computação. Isso pode ter confundido o desenvolvimento teórico da OC, porque teorias relacionadas a “informação” versus teorias relacionadas a “documentos” estão relacionadas a tipos de perspectivas bastante diferentes. Como conceito teórico, “informação” tende a mover a LIS e a OC em direção a teorias sobre controle, feedback, codificação e ruído na transmissão de mensagens, enquanto “documento” tende a mover a LIS em direção a teorias sobre significado, linguagem, conhecimento, epistemologia e sociologia. Portanto, em LIS pode haver todo um conflito paradigmático oculto nessas palavras.

Este terceiro estágio no desenvolvimento da OC como um campo foi influenciado por abordagens experimentais em que revocação e precisão são medidas bem conhecidas, pelo uso extensivo de modelos estatísticos das propriedades das representações de documentos, por abordagens que tentam automatizar OC usando Inteligência Artificial (IA) e sistemas especialistas, aplicando técnicas de processamento de linguagem natural (PNL) e similares. Um desdobramento desse estágio é também a visão cognitiva que tenta modelar as funções cognitivas dos usuários nos computadores.

O estabelecimento de serviços de resumo baseados em computador, como o Chemical Abstracts e o MEDLINE na década de 1960, é uma conquista importante neste estágio. O

desenvolvimento de recuperação de texto livre e baseada em descritores (principalmente com base em títulos e resumos), lógica booleana, acesso a assuntos específicos de campo, medição de revocação e precisão e outras inovações foram desenvolvimentos extremamente importantes na recuperação de documentos. A recuperação de informações como uma tradição de pesquisa começou com os experimentos ASTIA e Cranfield na década de 1950, e os experimentos de texto completo do TREC de hoje podem ser vistos como uma continuação dessa tradição. Este terceiro estágio melhorou os serviços de informação e os esforços de pesquisa em CI da maneira mais importante. Em minha opinião, essas inovações tecnológicas e os objetivos que tentam resolver definem a área central do que se denomina Ciência da Informação.

A tecnologia da computação tornou possível o uso de muitos tipos de Pontos de Acesso de Assunto (SAPs), tanto dos tipos tradicionais produzidos por especialistas em informação quanto de palavras dos próprios documentos (por exemplo, títulos e resumos). Isso removeu o monopólio dos bibliotecários / especialistas em informação sobre o acesso ao assunto e estabeleceu uma competição direta entre SAPs produzidos por diferentes agências.

Uma filosofia subjacente neste estágio costuma ser que o comprimento do registro pesquisável em si era o parâmetro mais importante na recuperação (Lancaster, 1998, pp.6-8). Os SAPs eram frequentemente vistos como meras “condensações semânticas” dos textos representados (implicando que o objetivo final era a representação do texto completo e nada mais). A pesquisa foi dominada por metodologias quantitativas, enquanto poucas pesquisas sobre diferenças qualitativas (semânticas ou significados) entre diferentes tipos de SAPs foram estabelecidas nesta fase. A filosofia subjacente foi empirista em mais de um sentido desta palavra. Foi empirista antes de mais nada em sua tentativa de medir a eficiência dos pontos de recuperação do assunto empiricamente, por exemplo, medindo "revocação" e "precisão". Foi também empirista em evitar classificações baseadas "metafisicamente", e se favorecer de SAP "atomistas" como o sistema Uniterm concebido por Mortimer Taube em 1951 e sistemas semelhantes, dependendo de palavras específicas do próprio documento.

Uma tendência associada neste estágio ou abordagem foi a tentativa de formalizar e automatizar a recuperação e eliminar a interpretação humana e a análise do assunto. Devemos distinguir entre a pressão financeira para automatizar os sistemas práticos de um lado e a avaliação científica do desempenho de vários aspectos dos sistemas de recuperação de base humana e mecanizada do outro. É uma meta legítima e altamente desejável reduzir custos e melhorar a eficiência dos sistemas de informação. A pesquisa básica deve, no entanto, iluminar os pontos fortes e as desvantagens básicas em diferentes abordagens, e não ser cegada pela pressão para usar soluções automatizadas ou baratas. Por causa de tais tendências, abordagens importantes relacionadas à interpretação e aos aspectos qualitativos têm sido muito negligenciadas nesta fase, e a pesquisa não acumulou um corpo de conhecimento tão satisfatório quanto o necessário.

d. Recuperação baseada em citações e OC (1963-)

A introdução do *Science Citation Index* por Eugene Garfield em 1963 marca o quarto estágio importante no desenvolvimento de SAPs. A possibilidade de recuperar documentos de acordo com as citações que recebem representa uma verdadeira inovação em RI, e essa técnica é capaz de complementar todas as formas de recuperação baseada em termos de novas *formas* muito importantes e qualitativas. Essa inovação também contribuiu para a pesquisa dos motivos de citação de outros documentos, nos padrões sociológicos na citação, no papel relativo dos termos e referências como SAPs e nas relações semânticas entre artigos citados e citações. No geral, esta etapa trouxe o conceito de *redes de documentos* para o foco da OC.

Desse modo, a recuperação baseada em citações mudou nossa compreensão não apenas da relação do assunto, mas também do conceito de assunto e do objetivo fundamental da própria RI. Uma vez que pode ser relevante citar artigos, que não têm palavras em comum com os artigos citados (ou nenhuma relação semântica simples, como termos mais restritos, termos mais amplos, sinônimos, etc.), as concepções ingênuas de relacionamento ou assunto não podem mais persistir. As relações semânticas podem ser implícitas ou latentes. As relações semânticas na ciência são determinadas por avanços teóricos, que podem mudar completamente a descrição verbal dos fenômenos de pesquisa. Esta é a razão pela qual os padrões estatísticos no vocabulário às vezes podem revelar-se uma medida menos eficiente de relacionamento entre assuntos do que os padrões nas citações.

O comportamento da citação é extremamente importante como objeto teórico porque o objetivo da RI é fornecer as referências que sejam úteis na resolução de um problema específico. Um artigo científico é uma documentação da solução de um problema de pesquisa específico. O problema é formulado no artigo, e este problema determinou que tipo de informação era necessária ao autor para contribuir com este problema. Com base na necessidade de informação, as informações foram buscadas e selecionadas, e os documentos efetivamente utilizados são finalmente citados no artigo. Cada um dos milhares de artigos produzidos semanalmente é, de certa forma, um estudo de caso em RI. Não apenas todo artigo apresenta um problema definitivo de RI, mas a lista de referências fornecida pelo autor é a chave de como aquela pessoa em particular resolveu o problema. Assim, é possível comparar as teorias de RI com o modo como elas correspondem aos documentos reais citados! Na filosofia da ciência tradicional (positivista), os estudos devem ser capazes de “prever” comportamentos ou eventos futuros. Em nosso caso, esse princípio implicaria que teorias e modelos de RI deveriam ser capazes de prever quais citações aparecerão em um único artigo. A maioria das pesquisas sobre relevância e RI parece ter esquecido esse fato. Pelo que sabemos, parece extremamente improvável que um algoritmo seja capaz de selecionar referências de bancos de dados eletrônicos e terminar com apenas o conjunto de referências representado em um determinado artigo. Deste ponto de vista, as teorias de RI parecem ingênuas e irrealistas e o objetivo da previsão parece estar errado. Um estudo mais detalhado do comportamento de citação pode iluminar os problemas reais de RI: Os documentos citados não são simplesmente um conjunto de documentos que compartilham um conjunto fixo de atributos que não são representados nos itens não selecionados. Documentos que são semelhantes do ponto de vista dos algoritmos de recuperação não precisam ser cocitados, enquanto documentos que não são semelhantes são frequentemente cocitados. Algoritmos de recuperação comuns e práticas de citação parecem simplesmente refletir diferentes teorias sobre a relação com o assunto.

Uma vez que os autores podem citar outros artigos a fim de lisonjear ou impressionar, a previsão de quais referências um determinado autor selecionará no final para um determinado artigo não pode ser usada como um critério válido em RI. Os critérios para RI não devem ser baseados em motivos psicossociais, mas em princípios epistemológicos para o avanço do conhecimento público! Dessa forma, nossa visão dos índices de citação mudou, de uma maneira muito profunda, não apenas os métodos de RI, mas também o próprio conceito de relacionamento do assunto e o objetivo básico de recuperar informações. Não podemos mais considerar a previsão do uso individual como o critério ideal para RI, nem podemos considerar a RI como uma técnica livre de valores. Em vez disso, nós temos que enfrentar a percepção de que os objetivos da RI estão profundamente enraizados em normas epistemológicas para o que deve ser considerado como boa ciência e bom comportamento de citação.

Os pioneiros na integração de métodos bibliométricos com os métodos mais tradicionais de OC são Mike M. Kessler (1965), Miranda Lee Pao e Dennis B. Worthen (1989), Pao (1993) e Lorna K. Rees- Potter (1987, 1989), 1991). Também vejo minha própria pesquisa como pioneira a esse respeito, tanto teórica quanto empiricamente (por exemplo, Hjørland, 1992, 1993, 1998a + b, 2002b). Posteriormente, também foi usada por, entre outros, estudantes de pesquisa de Peter Ingwersen (por exemplo, Schneider & Borlund, 2002).

e. Texto completo, hipertexto e Internet (principalmente 1990)

A recuperação de texto completo marca a quinta e última etapa no desenvolvimento de SAPs. Até este ponto, os limites de espaço eram uma restrição importante no desenvolvimento de sistemas de acesso de assunto porque o comprimento do registro em si é um parâmetro importante na recuperação. Nesse estágio, cada palavra ou combinação possível de palavras em documentos de texto completo é um SAP potencial, assim como todo tipo imaginável de informação de valor agregado fornecida por autores, leitores ou intermediários. Dadas as representações de texto completo, o primeiro problema teórico importante a surgir é se *algum* tipo de informação de valor agregado é necessário. As informações extras fornecidas pela resumo e indexação - pelo menos em princípio - podem aumentar a recuperação e / ou precisão? Se este não for o caso, então parece que chegamos ao fim da linha, que nenhuma contribuição adicional de pesquisa ou prática em CI é mais necessária?

A resposta a essa pergunta está intimamente ligada a visões teóricas sobre o conceito de assunto. Na visão de Claus Poulsen (1994), assunto é algo que se expressa na literatura de forma transparente e autoevidente. Ao definir os assuntos dessa forma, é impossível até mesmo perguntar se um determinado texto sempre retrata a representação ótima de si mesmo. Ao definir os assuntos como potencialidades informativas ou epistemológicas, Hjørland (1992, 1997) estabelece a possibilidade de que os documentos possam estar implícitos ou mesmo errados sobre seu próprio assunto e, portanto, ainda existir a necessidade de profissionais da informação. Para dar um exemplo extremo: um documento sobre judeus escrito por um autor nazista não deve apenas ser indexado como sendo sobre judeus, como afirma ser. É importante tornar a visão nazista visível na análise do assunto (e, por exemplo, indexá-la como propaganda nazista sobre os judeus). Os assuntos não são objetivamente “dados”, mas são influenciados por visões mais amplas que são importantes para o buscador de informações conhecer e, portanto, idealmente deveriam fazer parte da análise do assunto. Se esta é ou não uma solução prática, econômica e realista é outra questão que deve ser esclarecida avaliando os sistemas de acesso de assuntos específicos.

A introdução de bancos de dados de texto completo enfatizou os problemas teóricos relativos ao assunto e significado e também tornou os estudos de composição altamente relevantes para CI. Nesta fase, os aspectos qualitativos tornam-se mais importantes do que nunca. Não se trata apenas do problema de fazer algoritmos eficientes para RI, mas também do problema de identificar os valores e objetivos subjacentes aos quais esses algoritmos irão servir (ver Hjørland, 2003b).

*

A descrição acima dos cinco estágios de desenvolvimento orientados à tecnologia em OC fornece, em minha opinião, uma imagem clara do que OC significa com LIS. Ao ver esses cinco estágios juntos, temos uma imagem bastante bem definida do tópico geral com o qual estamos lidando. Por trás dessas diferentes tradições está um certo objetivo geral que pode ser expresso como a otimização de OC em bibliotecas, bancos de dados, obras de referência e na Internet. É importante ter em mente esse objetivo porque é ele que confere a OC na LIS sua importância para a prática.

É importante perceber, no entanto, que o desenvolvimento foi muito impulsionado pela tecnologia. Esta não é uma circunstância satisfatória para uma disciplina com a ambição de ser uma ciência e de produzir teorias e princípios gerais. Embora esses cinco estágios definam um tópico e um objetivo comum, eles não definem um fundo acumulado de descobertas, teorias e princípios. Pelo contrário, muitas vezes existem visões conflitantes latentes entre esses estágios ou tradições. Essas tradições têm sido apenas em uma extensão muito limitada na forma de um diálogo. É por isso que os conceitos gerais para pensar sobre a OC estão ausentes ou pelo menos fortemente subdesenvolvidos. Nas próximas seções, examinaremos alguns desses conceitos, teorias e princípios que parecem mais importantes para essa tarefa.

O objetivo da LIS é escrever uma história de seu desenvolvimento teórico *abstraido* das tecnologias concretas nas quais seus princípios foram estudados. Isso é difícil porque a base de conhecimento geral não está bem estabelecida. Além disso, a tendência é que a biblioteca e a ciência da informação usem passivamente a tecnologia, sem contribuir muito para seu desenvolvimento. Para que a LIS seja capaz de contribuir com conhecimento valioso, seu foco deve ser abstraido de tecnologias concretas.

3 Tratamento da informação na divisão social do trabalho: A Organização do Conhecimento Social e Intelectual.

Hjørland (2002d) analisou a definição oficial de ciência da informação dada pela *American Society for Information Science and Technology* (ASIST), que afirma:

“A ciência da informação preocupa-se com a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação, com particular ênfase nas aplicações das tecnologias modernas nessas áreas”. (Borko, 1968; Griffith, 1980).

Esta definição não discute como outras profissões além dos cientistas da informação estão preocupadas com “a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação”. Na verdade, sabemos que muitas profissões podem alegar estar igualmente preocupadas. Ao negligenciar essa questão, a definição falha em definir ciência da informação, uma vez que uma definição de CI deve especificar o papel especial dos profissionais da informação no estudo ou tratamento da informação.

Os astrônomos, por exemplo, podem ser vistos como especialistas que identificam, processam e interpretam informações do universo. Os astrônomos “lêem” a natureza e os livros, mas a natureza é geralmente considerada a informação chave nas ciências. Como produtos de suas atividades, eles podem publicar suas descobertas empíricas e teóricas. A profissão de bibliotecário, documentação e informação, embora interessada em todos os tipos de documentos, tem um interesse central e especialização relacionados à comunicação de documentos publicados. Os cientistas da informação *não* são especialistas em interpretar informações das estrelas, mas são, em sua maioria, especialistas em informações documentadas por, como exemplo, astrônomos (tal como, indexação e recuperação de documentos astronômicos). Neste exemplo, as informações foram definidas em um sentido mais amplo do que normalmente está implícito na ciência da informação.

Assim como se pode dizer que os astrônomos lidam com informação profissionalmente, pode-se dizer que todas as outras profissões o fazem também. Editores, pesquisadores, historiadores, advogados, professores e qualquer outra pessoa podem ser considerados profissionais no manuseio de informações de uma forma ou de outra. O papel

dos especialistas em informação pode ser relativamente claro quando o grupo-alvo é, por exemplo, astrônomos: Os especialistas em informação são especialistas em formas de publicações, bancos de dados, ferramentas de referência e assim por diante. No caso de, por exemplo, historiadores e advogados, os limites são muito menos claros porque as informações-chave que essas profissões buscam, interpretam e usam estão contidas em publicações e outros documentos. O historiador, não o bibliotecário ou o especialista em informação, é o especialista em buscar, organizar, interpretar e utilizar os documentos (em sua maioria não publicados) necessários em seu trabalho profissional, mas um profissional da informação é mais profissional em relação a alguns problemas específicos, como bancos de dados, catalogação e assim por diante. Uma diferença básica no conhecimento sobre as fontes de informação é normalmente que o especialista do assunto começa de um ponto estreito e trabalha de baixo para cima em fontes de informação mais gerais, enquanto o especialista em informação começa com uma visão geral ampla e trabalha de cima para baixo para fontes de informação mais específicas. Desta forma, suas competências são complementares.

O foco especial da LIS está no conhecimento documentado produzido por seres humanos em algum tipo de documento de uso potencial para outros seres humanos. A luz das estrelas não é informação para a comunidade LIS, mas sim informação astronômica produzida e usada pelos astrônomos. Essa distinção pode parecer sutil, mas é importante para construir uma base teórica sólida para a OC.

A OC envolve dois tipos de organização: a organização intelectual do conhecimento também pode ser denominada organização cognitiva do conhecimento. Esta é basicamente a organização do conhecimento em conceitos, sistemas conceituais e teorias. Se usarmos o sistema periódico da química ou a taxonomia zoológica da biologia como a base dos sistemas de indexação, estamos usando sistemas intelectuais ou “cognitivos” de OC. A organização social do conhecimento, por outro lado, é basicamente as organizações em profissões, ofícios e disciplinas. Se nos referirmos a disciplinas em nossas representações de conhecimento, estamos usando sistemas “sociais” de OC. Deve-se mencionar que este princípio de organização social tem sido o básico na classificação tradicional de bibliotecas, como a Classificação Decimal de Dewey (CDD):

Um trabalho sobre a água pode ser classificado em muitas disciplinas, como metafísica, religião, economia, comércio, física, química, geologia, oceanografia, meteorologia e história.

Nenhuma outra característica da CDD é mais básica do que esta: ela espalha os assuntos por disciplina (Dewey, 1979, p. Xxxi).

Há muito pouco debate sério sobre esse princípio na literatura, apoiando a visão de que muitos bibliotecários e especialistas em informação parecem desconfortáveis com ele. Quer este princípio seja aprovado ou não, deve ficar claro, entretanto, que a organização social do conhecimento nas disciplinas é muito importante para a OC e para otimizar a busca de informações. Quando tiver sido estabelecido a qual disciplina pertence uma determinada consulta, a parte mais difícil da tarefa de recuperação pode muito bem ter sido concluída. A organização social do conhecimento é de fato importante para estudar, e além de algumas pesquisas em bibliometria (ver Hargens, 2000), este estudo mal começou na LIS, apesar do significado expresso na citação dada acima. Hoje, a base “empírica” para classificações como CDD é principalmente a literatura classificada pelo sistema. Uma forma de expandir a dimensão social poderia ser estudar o desenvolvimento das ocupações na sociedade, por exemplo, conforme exibido no *Standard Occupational Classification System* (U.S. Department of Labor, Bureau of Labor Statistics, 2002).

É claro que há diferentes teorias ou estruturas conceituais sobre a organização intelectual e social do conhecimento (bem como sobre sua inter-relação). Uma visão vê os conceitos científicos, teorias e campos como refletindo uma realidade neutra e objetiva. Isso pode ser denominado *ciência como uma metáfora do espelho* e está relacionado ao racionalismo. Os cientistas podem se organizar de acordo com esses campos pré-estabelecidos. Basicamente, entretanto, a organização intelectual não está conectada - muito menos formada - à divisão social do trabalho. Outra visão pode ver os conceitos científicos, teorias e campos como ferramentas úteis construídas para que os seres humanos possam acomodar as demandas da vida. Esta é a *ciência como metáfora do mapa*, que está relacionada ao pragmatismo.⁵ Mesmo que um mapa seja um reflexo ou representação de uma realidade, um mapa ainda não é um espelho. Um mapa é antes de tudo uma ferramenta para certas atividades humanas. O tipo de atividades e interesses que o mapa atenderá tem uma grande influência em como o mapa é feito. De uma perspectiva pragmática, a organização intelectual do conhecimento está profundamente enraizada e conectada à organização social do conhecimento.

Uma visão influente hoje é o *construtivismo social*. Essa visão está relacionada à visão pragmática apresentada acima. Frequentemente, porém, o construtivismo social e o pragmatismo se opõem a tipos de realismo como o realismo científico. Um dos filósofos pragmáticos modernos é Richard M. Rorty. De acordo com Rorty, o realismo científico e o pragmatismo são duas visões que não podem ser combinadas; o pragmatismo é visto como uma posição antirrealista. De acordo com outros filósofos (por exemplo, Dewey, 1929; Ellis, 1990), o realismo deve de fato ser baseado no pragmatismo. Isso corresponde à minha própria visão, que pode, portanto, ser denominada *realismo pragmático*. Os conceitos e campos científicos tendem a representar partes da realidade de uma forma funcional para a atividade humana. Na medida em que o construtivismo social se opõe ao realismo, discordo dessa visão. Deve-se evitar a armadilha do reducionismo sociológico. No entanto, descobri que muitas pesquisas feitas sob a bandeira do construtivismo social são profundamente relevantes para a compreensão da estrutura de muitos campos do conhecimento. Por exemplo, para iluminar a classificação da doença mental, é relevante fazer pesquisas experimentais e fazer "análise do discurso". A pesquisa experimental pode mostrar que a esquizofrenia não é um conceito etiológico frutífero porque todas as principais psicoses podem ter uma base genética comum (ver Kringlen, 1994). A "análise do discurso", por outro lado, pode revelar o fato de que a maioria de nossas teorias e conceitos neste campo são "construções sociais" influenciadas por interesses específicos. Em particular, pode chamar nossa atenção para diferentes experiências culturais e evitar tendências a explicações universalistas. Desta forma, pode abrir-se a classificações mais adequadas para um diálogo entre diferentes perspectivas. No final, nossos conceitos científicos e classificações (por exemplo, "esquizofrenia" versus "Einheitspsychose" – "psicose unitária") são ferramentas que devem ser avaliadas de acordo com o trabalho que devem fazer para nós. Como tal, são partes de teorias que foram desenvolvidas para nos fazer agir em relação a problemas específicos.

A organização do conhecimento em tradições, ideologias e paradigmas pode ser vista como a combinação de conceitos entre o intelectual e a organização social. Eles são organizações cognitivas baseadas em influências sociais.

4 Documentos e a perspectiva sociológica de produção de conhecimento, intermediação e consumo de conhecimento.

É importante para a ciência da informação fornecer modelos de atores, instituições e serviços de informação em diferentes comunidades de discurso. Os atores são produtores de conhecimento, intermediários e usuários de conhecimento. As instituições são instituições de pesquisa, editoras, bibliotecas e assim por diante. Os serviços de informação podem ser classificados em serviços primários (por exemplo, editoras e periódicos), serviços secundários (como exemplo, bancos de dados bibliográficos) e serviços terciários (por exemplo, enciclopédias profissionais e revisões de literatura). Achei o UNISIST (1971) útil. A Figura 1 abaixo é uma versão modificada do modelo UNISIST de Trine Fjordback Søndergaard, Jack Andersen, & Birger Hjørland (2003, p. 303), no qual a elipse pontilhada simboliza um domínio de conhecimento. A Figura 2 mostra o mesmo modelo (não publicado anteriormente) em que os recursos da Internet são destacados como integrados às fontes impressas.

Neste ponto, quero enfatizar que tais modelos fornecem uma perspectiva específica sobre a organização do conhecimento. Ao apresentar os atores e as divisões de trabalho entre diferentes organizações e serviços, esses modelos são sociológicos por natureza. Eles nos ajudam a compreender as diferentes funções dos tipos de literatura e nos permitem estabelecer uma tipologia significativa da literatura acadêmica, incluindo a bem conhecida diferenciação entre literatura primária, secundária e terciária.

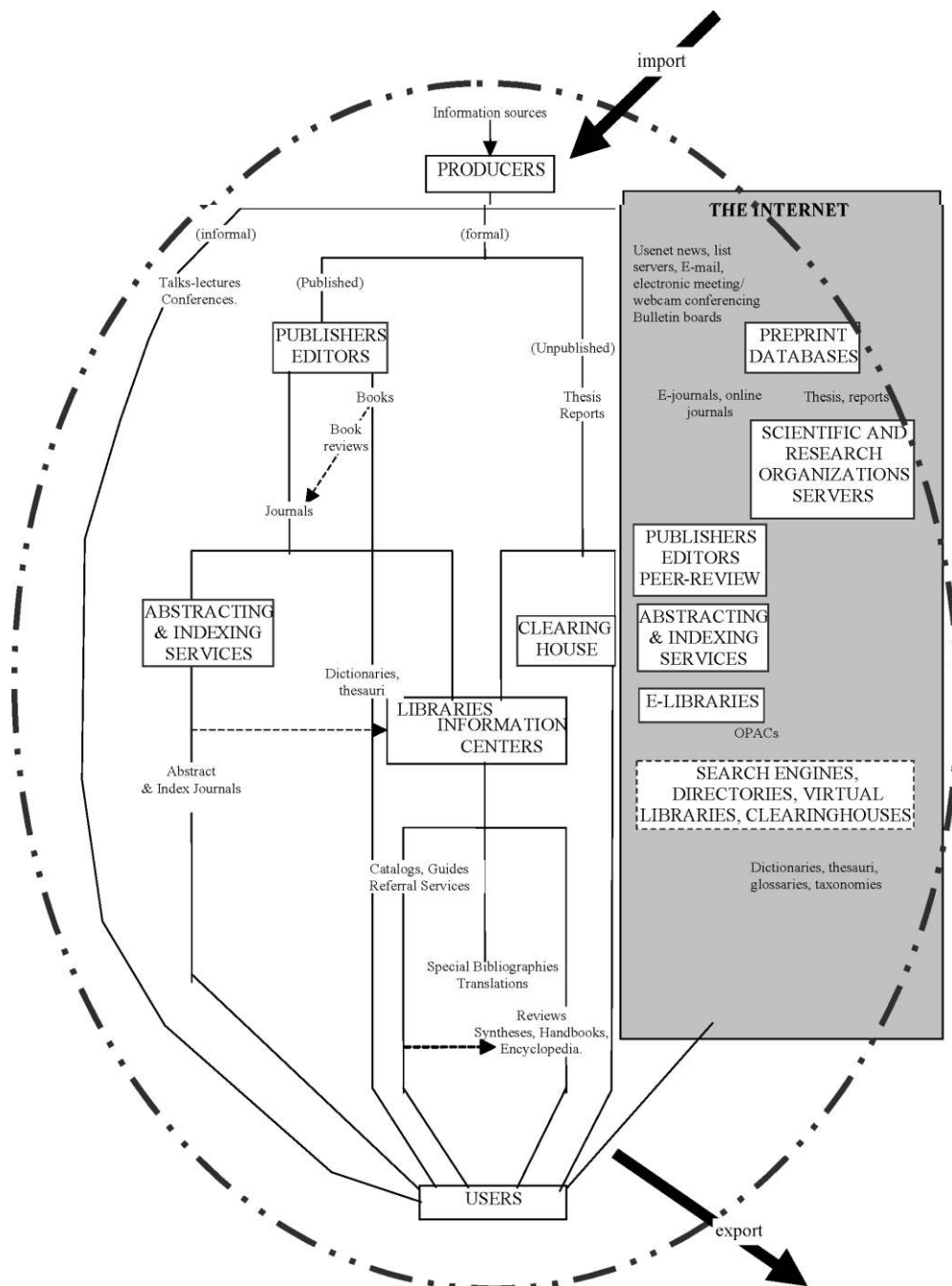


Figura 1: O modelo UNISIST revisado modificado para a abordagem analítica de domínio

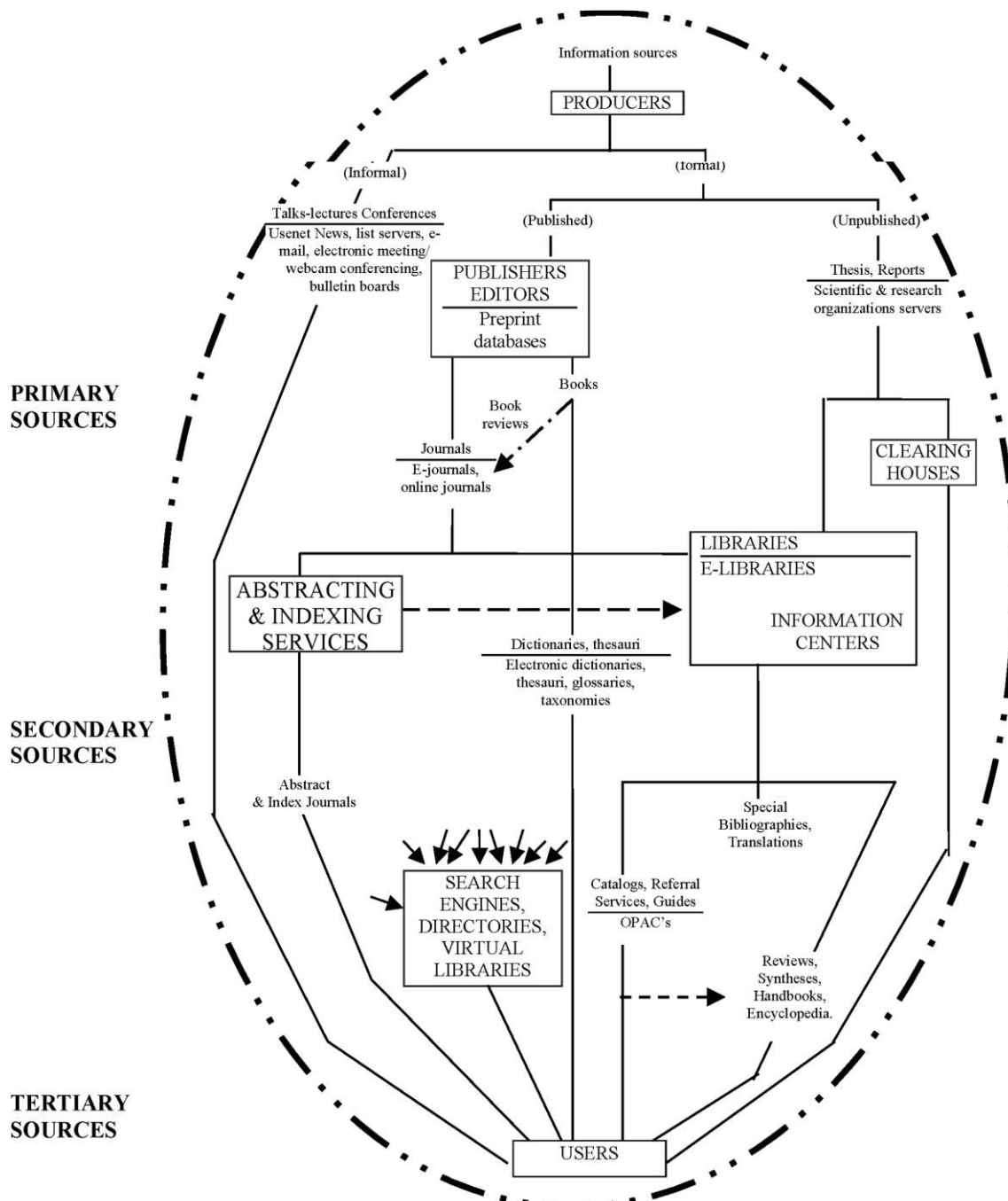


Figura 2: O modelo UNISIST revisado integrando recursos impressos e da Internet e modificado de acordo com a abordagem analítica de domínio

5 Pontos de acesso do assunto em recuperação eletrônica.

Na recuperação eletrônica, palavras, frases e textos são retirados de contextos e mesclados. Palavras e símbolos de muitos periódicos diferentes são mesclados; palavras de diferentes seções e partes de textos são mescladas e assim por diante. As palavras são tiradas de seu contexto, um contexto que contribui para seu significado. Usando tecnologia avançada, é possível restabelecer partes desse contexto perdido (por exemplo, por meio de operadores de proximidade). Qualquer sequência de caracteres e qualquer símbolo dos próprios textos ou das chamadas informações de valor agregado tornam-se um possível ponto de acesso ao assunto. A arte da ciência da recuperação da informação e da OC é utilizar todas essas

possibilidades de maneira otimizada e agregar mais valor à representação e recuperação de documentos.

O que devemos enfatizar em CI é estudar as maneiras pelas quais palavras e símbolos recebem significado em seus contextos específicos. São importantes a maneira como diferentes disciplinas constroem seus significados, a maneira como a composição do documento fornece significado a palavras e símbolos e a maneira como diferentes vocabulários controlados constroem significados. Se uma palavra for empregada, por exemplo, no título, nos resumos, na seção de metodologia ou na conclusão, podemos (às vezes) ser capazes de atribuir diferentes prioridades e significados à palavra, e isso pode nos ajudar a dar diferentes avaliações probabilísticas sobre se os documentos devem ser recuperados ou não. Pessoas, que conhecem certos bancos de dados muito bem, incluindo a literatura do assunto e a linguagem do assunto que cobrem, provavelmente empregam esse tipo de conhecimento de maneira tácita. O trabalho dos estudos de informação é ajudar a explicar os princípios subjacentes. Assim como as pessoas comuns podem falar uma língua e usar a gramática, os especialistas em linguística precisam explicar o que os usuários competentes fazem.

Em Birger Hjørland e Lykke Kyllesbech Nielsen (2001), resumimos a pesquisa sobre o valor informativo relativo de diferentes tipos de pontos de acesso do assunto (SAPs). É importante perceber que o valor relativo de, por exemplo, termos de uma pesquisa de texto versus citação (com base nas referências bibliográficas) não é constante. Varia de acordo com as normas disciplinares, entre outras coisas. Isso implica que é importante descobrir essas normas disciplinares diferentes: Em nossa revisão mencionada acima, tentamos integrar esse conhecimento de muitas fontes diferentes. Nossa abordagem implica que a rica flora de documentos e domínios deve ser investigada pela ciência da informação. Isso está em contraste com as abordagens prevaletentes “universalistas”, “reducionistas” ou “individualistas”. O objetivo dos estudos é explicar e representar os significados implícitos que se perdem na fusão.

As três principais abordagens para RI e OC são:

- a) RI tradicional com base nas frequências dos termos nos próprios textos e na base de dados total.
- b) Organização e recuperação de bases de citações com base em redes de citações entre documentos.
- c) Classificações tradicionais da biblioteca com base em divisões disciplinares e critérios disciplinares. Esses critérios disciplinares são, por princípio, externos aos próprios documentos e NÃO são derivados de bancos de dados de texto completo ou redes hiperligadas. Você não pode, por exemplo, construir um mapa da Espanha a partir de uma coleção de documentos sobre a Espanha, quer use a análise de cocitação ou qualquer outra técnica. Os mapas são produzidos por geógrafos e suas estruturas conceituais são posteriormente aplicadas na classificação da biblioteca.

Todas essas abordagens costumam ser usadas simultaneamente e os registros bibliográficos costumam conter uma mistura de todas elas. Devemos desenvolver mais conhecimento (explícito) sobre seus pontos fortes e fracos relativos para diferentes tipos de consultas.

A OC envolve dois tipos de decisões: Fazer um sistema de classificação (ou um tesouro, etc.) e decidir onde classificar um determinado documento nesse sistema. Ambas as decisões estão relacionadas e envolvem implicações epistemológicas. Ambas as decisões são feitas no final, em última análise com base no que é considerado relevante: quais classes são

relevantes? Que tipos de conexões entre as classes são relevantes? Que tipos de documentos são relevantes para determinados usuários ou propósitos para reunir em uma determinada classe?

Diferentes tipos de SAPs têm diferentes valores informativos na recuperação de informações. Os SAPs são mais ou menos funcionais ao revelar “o assunto” dos documentos que representam. É importante perceber, entretanto, que quais documentos são relevantes para uma determinada consulta depende, entre outras coisas, da perspectiva teórica em que a consulta está inserida. Uma descrição do “assunto” não é uma atividade neutra e objetiva, mas é influenciada por diferentes pontos de vista teóricos e interesses. A consequência dessa visão é que “o próprio assunto” (algo que os SAPs supostamente identificam) não pode ser considerado objetivo no sentido positivista. Com base em tais considerações, Hjørland (1992) definiu o assunto de um documento como as *potencialidades* epistemológicas (ou informativas) do documento. Um documento traz potencialidades, e o trabalho dos indexadores (e outros SAPs) é identificar essas potencialidades. Isso é facilmente compreendido se compararmos com os índices de citação. Pode haver muitos motivos para citar um determinado documento, mas alguns desses motivos podem ser bastante variados. Para o grupo de autores que citam um determinado documento pelo mesmo motivo, este documento tem a mesma função e o mesmo “assunto”. A indexação é um processo a serviço de ajudar futuros autores a identificar documentos a serem citados. Isso é feito identificando exatamente os atributos que o tornam citável. As conexões entre os artigos citados e os citantes são conexões de relevância, e também são (ou deveriam ser) as conexões entre os termos de indexação e os documentos indexados.

6 Semiótica Social e Teoria da Atividade: A Natureza Teleológica da OC

A visão estruturalista de como nossos conceitos são formados por nossas linguagens é mostrada na Figura 3 abaixo. Mostra que não existe uma relação direta entre os significados em diferentes línguas. Quando os indivíduos aprendem uma língua, aprendem os conceitos dessa língua e, conseqüentemente, classificam o mundo da maneira que é dada por aquela língua específica. Por exemplo, alemães e dinamarqueses classificam “árvore” de maneiras diferentes. Por implicação, as linguagens podem afetar a maneira como conceitualizamos o mundo. De acordo com Louis Hjelmslev (1943), cada linguagem coloca fronteiras *arbitrárias* na realidade, enquanto a teoria da atividade (TA) descobre que nossos sistemas simbólicos tendem a capturar aspectos funcionais ou possibilidades nas coisas que percebemos (ver Albrechtsen, Andersen, Bødker, & Pejtersen, 2001).

English	*German	*Danish	*French	Italian	Spanish
Tree	Baum	Træ	Arbre	albero	Árbol
Wood	Holz		Bois	legno	Leña
Woods	Wald	Skov		bosco	Bosque
Forest			Forêt	foresta	Selva

Figura 3: relatividade cultural nos significados das palavras

* Apresentado originalmente pelo linguista estrutural dinamarquês Louis Hjelmslev (1943). Ampliado por informações de Buckley (2001)

A figura de Hjelmslev pode ser criticada; por exemplo, foi mencionado que ele omite o termo dinamarquês “lund” (pequena floresta) e que esta omissão pode enfraquecer este exemplo particular. Neste ponto, entretanto, devemos tomar esta afirmação estruturalista básica como certa: que determinadas línguas afetam nossas estruturas semânticas e, portanto, nossa classificação do mundo.

Para atividades informativas humanas, a perspectiva adequada do significado de “significado” é muito importante. Este é um conceito difícil para o qual a teoria da atividade fornece uma compreensão frutífera. A produção de livros, textos e outros documentos é um desenvolvimento especial nas culturas letradas. Documentos são ferramentas com tipos específicos de valores funcionais nessas culturas. A visão da TA e das teorias semióticas sociais⁶ é que significados, signos e documentos são desenvolvidos para funcionar em relação a práticas padronizadas nas comunidades. Usamos, por exemplo, a Bíblia e o Livro de Hinos em nossas práticas religiosas padronizadas, livros em nossas práticas de ensino padronizadas, livros jurídicos em práticas legais padronizadas e assim por diante.

Quando nomeamos algo, facilitamos o uso desse objeto. Ao nomear igrejas (ou livros sobre igrejas) “atrações turísticas”, facilitamos um determinado uso de igrejas (ou documentos). Realizamos um ato por meio da linguagem. Os nomes próprios (ou SAPs) sempre facilitarão alguns usos de um documento e tornarão outros usos relativamente mais difíceis. Nenhuma terminologia, tecnologia ou tipo de SAP pode ser neutro.

Conceitos e documentos têm valores funcionais mais ou menos estáveis em relação a tais práticas padronizadas. É claro que sempre há pontos de vista diferentes sobre se essas práticas padronizadas devem ser alteradas ou permanecer inalteradas, e sempre há diferentes tipos de mudanças possíveis dessas práticas. Sempre há desenvolvimentos dinâmicos de significado. A causa mais importante disso é o propósito para o qual usamos os conceitos, que está conectado às práticas que queremos mudar.

Frequentemente, o desenvolvimento científico e tecnológico muda as práticas padronizadas de uma forma bastante incontroversa. Em outros casos, no entanto, as mudanças nas práticas estão relacionadas a diferentes interesses políticos, a diferentes teorias ou “paradigmas”. Diferentes paradigmas tendem a influenciar determinadas práticas de maneiras diferentes e, ao fazê-lo, também tendem a mudar nossos sistemas simbólicos, bem como nossa produção de documentos e a forma, conteúdo, significado e uso desses documentos. O estudo adequado de símbolos e documentos é, portanto, baseado no estudo das funções e interesses que esses documentos atendem.

Qualquer ferramenta dada (incluindo um signo, conceito, teoria, achado ou documento) traz potencialidades em relação a novas práticas e paradigmas sociais. Usuários de bibliotecas podem descobrir que documentos anteriormente negligenciados podem ser muito úteis para desenvolver uma nova prática social ou para defender uma prática existente. Ou eles podem descobrir que documentos conhecidos podem ser reinterpretados e usados de novas maneiras. Principalmente os tipos de documentos conhecidos como “clássicos” têm tais habilidades para serem reutilizados e reinterpretados por cada geração. Mas mesmo os documentos mais humildes podem acabar sendo valiosos. Isso, é claro, não implica relativismo em relação à relevância dos documentos: as ferramentas também se tornam rapidamente obsoletas e “apenas de interesse histórico”. É claro que também pode haver uma discrepância entre o valor potencial de alguns documentos e o custo de armazenamento, organização e uso.

O significado de qualquer signo é sua qualidade potencial para se referir a alguns objetos ou estados de coisas (Karpatschhof, 2000, p.197). Uma ferramenta é algo que tem um valor funcional para alguma (sub) cultura humana. Linguagens (e sublinguagens) também são ferramentas com valores funcionais. Em idiomas, existem termos para ferramentas. O

significado de uma palavra para um determinado tipo de ferramenta é o valor funcional de se referir a um determinado valor funcional, definindo a qualidade da ferramenta (Karpatschhof, 2000, p.197). Um martelo pode, por exemplo, ser denominado “martelo” (denotação) ou “arma do crime” (conotação). A palavra que usamos sobre uma ferramenta facilita um ou outro uso dela. As palavras podem ser usadas mais ou menos apropriadamente para um objeto específico em relação a uma dada tarefa, atividade ou discurso. Em outras palavras, usar palavras é um tipo de ato (atos verbais) frequentemente usado para realizar algo extraverbal. Os significados das palavras que usamos podem ser mais ou menos adequados para nossos propósitos e, nesses casos, tentamos desenvolver novas palavras ou mudar o significado de algumas palavras antigas. Essas mudanças de significados são visíveis no estudo de diferentes campos, tradições e paradigmas. É muito ingênuo e reducionista desconsiderar esses tipos de fatores de intermediação cultural na relação das pessoas com a informação. As tradições dominantes na ciência da informação e nas ciências comportamentais e cognitivas, no entanto, negligenciaram esses aspectos culturais e apenas tentaram estudar as relações humanas generalizadas para algo denominado "informação". Essa abordagem dominante pode ser amplamente denominada de “comportamental”, apesar das diferentes atitudes em relação às versões do behaviorismo. Nessa tradição, espera-se que as pessoas reajam a algo de uma forma específica e mecânica, sem considerar os significados culturalmente determinados e sem considerar os diferentes objetivos e valores nos significados e nos documentos. Isso, em minha opinião, gerou uma situação em que herdamos muito pouco conhecimento útil dessas áreas para fazer avançar nosso campo e nossa prática. Essa visão também é expressa por Robert de Beaugrande (1997) em linguística. “Eu certamente ficaria mais feliz se minhas descobertas tivessem se tornado menos serias ou preocupantes; mas só posso relatar o que de fato encontrei ” (§89).

Na visão do pragmatismo (e da Teoria da Atividade), as linguagens são ferramentas que são adaptadas culturalmente para atender às necessidades de seus usuários. As linguagens gerais podem ser vistas como adaptações que atendem às necessidades da maior parte das populações, enquanto as linguagens para fins especiais são adaptações que atendem às necessidades de grupos específicos, como químicos, advogados, músicos e assim por diante. Ambas as linguagens, (ou mais amplamente sistemas simbólicos) como também a mídia, documentos e os sistemas de informação, são teleológicas ou direcionadas a um objetivo. Eles são otimizados para realizar certas funções, enquanto ignoram relativamente outros propósitos e objetivos. Eles devem ser vistos como especializações na divisão social do trabalho. Isso implica que nenhum sistema de OC pode ser otimizado para fazer todos os tipos de tarefas igualmente bem. Embora os modernos sistemas de recuperação baseados em computador sejam muito flexíveis e pareçam eficientes para quase todas as tarefas, é importante considerar as limitações de cada tipo de sistema para diferentes tipos de tarefas e diferentes domínios. A forma como o conhecimento é organizado em sistemas de informação deve ser relevante para o propósito específico daquele sistema em particular. A relevância deve, em minha opinião, sempre ser considerada em relação a um objetivo. Birger Hjørland e Frank Sejer Christensen (2002) definiram relevância desta forma: “*Algo (A) é relevante para uma tarefa (T) se aumentar a probabilidade de cumprir a meta (G), que está implícita em T*” [p.964]. Nossa concepção de relevância está, portanto, conectada ao significado e às diferentes visões e tradições da sociedade. É uma visão básica na tradição pragmática que a natureza do conhecimento é cumprir objetivos para o organismo ou sistema que possui o conhecimento. Todos os tipos de conhecimento ou processos ou instituições de processamento de informação são basicamente teleológicos por natureza. Eles têm que cumprir certos objetivos. Seus critérios de relevância são construídos a partir desses objetivos, e as

pessoas aprendem esses critérios de relevância sendo socializadas e educadas dentro de um contexto e tradição específicos.

7 Paradigmas, epistemologias e OC

O conceito de "paradigma" tornou-se influente através do livro de Thomas Kuhn (1962, 1970, 1996). Kuhn, entretanto, não reconheceu as ciências sociais como ciências ou como paradigmáticas. Ele considerava as ciências sociais pré-paradigmáticas. Em sua opinião, um paradigma deve ser um conjunto de suposições *comuns*. Conforme afirmado por Dogan (2001):

Dentro de uma disciplina formal, várias teorias principais podem coabitar, mas só existe um paradigma quando uma teoria testável sozinha domina todas as outras teorias e é aceita por toda a comunidade científica. Quando Pasteur descobriu o micróbio, a teoria da geração espontânea entrou em colapso: o contágio tornou-se o novo paradigma. Nas ciências sociais, entretanto, vemos, na melhor das hipóteses, um confronto entre várias teorias não testáveis. Na maioria das vezes, não há nem mesmo um confronto, mas sim uma evitação mútua cuidadosa, um desprezo soberbo de todos os lados. Isso é relativamente fácil, devido ao tamanho das disciplinas científicas e sua divisão em escolas; é verdade para todos os países, grandes ou pequenos. (p. 11024)

No entanto, sua teoria tem sido muito influente nas ciências sociais, e o conceito de paradigma é frequentemente usado em relação às diferentes escolas, abordagens, sistemas e assim por diante. Assim, o conceito de paradigma costuma ser aplicado de forma mais ampla. "Paradigma" pode, na filosofia da ciência, significar suposições de fundo mais ou menos implícitas sobre o objeto de pesquisa, sobre métodos de pesquisa, sobre a utilidade da pesquisa e assim por diante. Em um sentido ainda mais amplo, o conceito pode ser usado com visões mais ou menos coerentes a respeito de uma prática coletiva, como ensino, serviço divino, administração da justiça, política e assim por diante. O pós-escrito de Kuhn de 1970 na segunda edição define, entre outras coisas, um paradigma como uma constelação de crenças, valores, técnicas e assim por diante compartilhados por membros de uma determinada comunidade.

Håkan Tjörnebohm, em 1974, (conforme citado em Andersen, 1999, p. 89), definiu paradigmas como sistemas de pressupostos básicos (explícitos ou implícitos) e ideais epistêmicos em disciplinas científicas. Um paradigma é uma estrutura superindividual de significado, que se forma e se reproduz na socialização disciplinar, no ensino e na comunicação científica. Os componentes dos paradigmas incluem:

- 1) Ideais e crenças sobre a ciência (objetivos epistêmicos, métodos e critérios na produção e avaliação dos resultados científicos dentro da disciplina);
- 2) Hipóteses de visão de mundo, pressupostos ontológicos básicos; e
- 3) Ideais relativos ao significado para a sociedade e cultura, para uso prático e para a iluminação.

"Paradigma" no sentido mais amplo da palavra é um conceito muito central para OC. Os paradigmas científicos não apenas determinam como estudamos a OC na ciência da informação (ver seções 9 e 10 abaixo); também é o caso que o conhecimento que vamos organizar já está conceituado e organizado segundo estruturas mais ou menos invisíveis,

determinadas por diferentes influências paradigmáticas. Isso é feito em muitas “camadas” em linguagens comuns, em linguagens de assunto, em formas documentais, em redes de citações de artigos, em redes informais e assim por diante. Diferentes paradigmas geralmente implicam em valores e objetivos diferentes. A OC em ciência da informação deve conhecer esses diferentes valores e ser capaz de cumprir os diferentes objetivos.

8 Conceitos e relações conceituais como unidades de OC

Os conceitos são normalmente vistos como unidades de pensamento e conhecimento. Isso implica que as unidades fundamentais da OC são as relações (semânticas) entre os conceitos. Se o conhecimento é definido como crença verdadeira justificada (como na tradição platônica), então o conhecimento real é difícil ou impossível de identificar e classificar. É mais frutífero falar de afirmações de conhecimento do que do próprio conhecimento. Falar de *afirmações de conhecimento* como as coisas representadas na literatura e a coisa a ser classificada é uma maneira mais cuidadosa de falar, e não há perda real com essa maneira de falar.⁷

A função básica dos conceitos é fornecer uma base para lidar com o mundo. Os conceitos fornecem fronteiras e classes em um mundo contínuo. “Azul” delimita alguns comprimentos de onda, “castelo” delimita alguns tipos de edifícios e “música” delimita alguns sons. Nossas teorias, conceitualizações e paradigmas tendem a classificar as coisas e o conhecimento sobre as coisas de acordo com os mesmos princípios básicos. Anders Ørom (2003, no prelo) demonstrou como, por exemplo, imagens em museus, a literatura sobre arte e as classificações usadas em bibliotecas e bibliografias tendem a ser influenciadas pelo mesmo paradigma (por exemplo, o paradigma estilístico, o paradigma histórico cultural, o paradigma semiótico, o paradigma histórico-social, o paradigma feminista etc.).

Um descritor ou uma classe representa um conceito⁸, e a recuperação da informação está essencialmente relacionada às relações semânticas entre consultas, representações de documentos e textos. A menor unidade são as relações entre dois conceitos. A teoria dos conceitos (significado, semântica) é, no entanto, provavelmente um dos campos de pesquisa mais difíceis e confusos da atualidade. É importante compreender as diferentes visões entre as teorias tradicionais baseadas no positivismo lógico e visões alternativas baseadas em teorias pragmáticas. Temmerman (1997) descreve algumas dessas diferenças.

Observe a distinção entre, por um lado, a *classificação* lógica e ontológica, que na [teoria da] Terminologia tradicional é considerada possível na mente sem considerar ou usar a linguagem e antes que a nomenclatura dos conceitos ocorra, e, por outro lado, a *categorização* que é resultado da interação entre a linguagem e a mente (p. 55).⁹

A importância dos conceitos foi reconhecida na comunidade OC. Dahlberg (1991) é um exemplo proeminente. A revista *Knowledge Organization* tem como subtítulo “International Journal”. Dedicado à Teoria Conceitual, Classificação, Indexação e Representação do Conhecimento”. Apesar disso, as contribuições sobre conceitos até agora têm sido muito limitadas em número, e o conceito de conceitos é muito melhor explorado em outros campos como filosofia, psicologia cognitiva e do desenvolvimento, inteligência artificial e linguística.

Diferentes teorias de conceitos estão profundamente ligadas às teorias epistemológicas. Empirismo, racionalismo, historicismo e pragmatismo têm suas diferentes concepções de conceitos.

<i>“Paradigmas”</i>	<i>Em filosofia</i>	<i>Em psicologia</i>
<i>Empirismo</i>	<p>Conceitos simples correspondem a sensações simples.</p> <p>Não há relações necessárias entre conceitos simples. Conceitos simples podem ser combinados com conceitos complexos. <i>Nominalismo</i>: conceitos gerais são nomes, que atribuímos a classes de coisas (generalizações empíricas).</p>	<p>Associacionismo clássico. Behaviourism. Conexionismo (neoassociacionismo ou processamento paralelo distribuído, que funciona com redes indutivas, auto-organizadas e "neurais" para o processamento de entrada sensorial).</p>
<i>Racionalismo</i>	<p>Simples são conceitos, que não podem ser definidos de maneira fecunda por outros conceitos. Eles não são experientes, mas inatos ou maturados. Conceitos complexos são definidos a partir de conceitos simples. Conceitos fundamentais são conceitos necessários para descrever um campo. Conceitos simples, complexos e fundamentais entram em certas relações mútuas necessárias. A diferenciação entre conceitos simples e complexos é absoluta (independente de domínio, interesses, pontos de vista etc.).</p>	<p>Inteligência Artificial Clássica. Cognitivismo. Trabalha com algoritmos dedutivos governados por regras para o processamento de dados sensoriais. Conceitos “clássicos” ou “aristotélicos” exaustivamente definidos por conjuntos de atributos necessários e suficientes.</p> <p>Teoria moderna (Rosch, Lakoff) Prototypetheory. Os conceitos são instâncias mais ou menos prototípicas das coisas.</p>
<i>Críticismo (Kant) (Racionalismo-empírico)</i>	<p>Os conceitos representam o conhecimento do mundo mediado por nossas formas de razão ou categorias como espaço, tempo, coisa e causa.</p>	<p>A “epistemologia genética” de Jean Piaget.</p> <p>“Os conceitos amadurecem no indivíduo. Eles crescem como os dentes”</p>
<i>Historicismo e hermenêutica</i>	<p>Os conceitos são formados em um processo histórico com base na pré-compreensão e na percepção holística. Existe uma circularidade entre a formação de conceitos simples e complexos. As relações entre conceitos simples e complexos são relativas em relação aos interesses. Tradições e comunidades sociais desempenham papéis importantes para a formação de conceitos.</p>	<p>Interpretação psicanalítica. Uma análise satisfatória das funções da cognição não pode apenas explorar a sensação, a memória e o pensamento isoladamente, mas deve envolver toda a pessoa e sua história de desenvolvimento tanto individual quanto coletivamente. Os conceitos são, portanto, formados pela influência de características pessoais, como sexo e classe social.</p>
<i>Pragmatismo e realismo crítico</i>	<p>Conhecimento e conceitos são formados pela atividade prática das pessoas em relação aos objetos da atividade.</p>	<p>Teoria da atividade: nossos conceitos não são determinados apenas por atributos em objetos físicos. Eles também têm “profundidade histórica”. Eles são formados nos contextos históricos dos objetos, que têm significado.</p>

Figura. 4: Conceitos básicos de "conceito"

Frank C. Keil descreve alguns desenvolvimentos muito importantes nas teorias sobre conceitos:

... uma coisa tem surgido cada vez mais. Sistemas de crenças coerentes, ou “teorias”, são essenciais para a compreensão da natureza dos conceitos e como eles se desenvolvem. Embora filósofos e alguns psicólogos de tendência mais clássica muitas vezes tenham argumentado sobre a interdependência de teorias e conceitos, em sua maioria os psicólogos cognitivos não o fizeram nos últimos 20 anos. Como consequência do trabalho feito na década de 1960 sobre aprendizagem de conceitos e a posterior resposta “anticlássica” a esse paradigma, tornou-se a norma falar sobre conceitos como consistindo em combinações de recursos ou como sendo o resultado de operações resumidas em exemplares e / ou dimensões. Embora as visões probabilísticas clássicas e mais recentes colidissem sobre a questão de se as características para conceitos naturais eram necessárias e suficientes, seu debate mascarou um conjunto compartilhado de suposições de que os significados dos conceitos poderiam ser totalmente descritos por listas de características e operações probabilísticas ou correlacionais simples nesses recursos.

À medida que os pesquisadores começaram a examinar os fenômenos associados aos conceitos mais de perto, no entanto, tornou-se cada vez mais claro a partir de várias linhas de evidências diferentes que algo estava faltando Parte do significado de um conceito pode ser a apreensão das relações teóricas que explicam sua estrutura interna tais relações sempre foram fundamentalmente importantes na filosofia da ciência. (Keil, 1989, p. 267).

A história de todas as ciências naturais documenta a descoberta de que certas entidades que compartilham propriedades imediatas, não obstante, pertencem a tipos diferentes. A biologia oferece muitos exemplos, como as descobertas de que golfinhos e baleias não são peixes, mas mamíferos, que o morcego não é uma espécie de pássaro, que a “cobra” de vidro é na verdade uma espécie de lagarto com apenas vestígios de membros abaixo de sua pele. No reino vegetal, descobriu-se, por exemplo, que alguns “vegetais” são realmente frutas e que algumas “folhas” não são realmente folhas. Do reino dos minerais e elementos vieram as descobertas, entre outras, que o mercúrio é um metal e que a água é um composto ...

Em quase todos esses casos, as descobertas seguem um curso semelhante. Certas entidades são inicialmente classificadas como membros de uma espécie porque compartilham muitas propriedades salientes com outros membros genuínos desse tipo e porque sua associação está de acordo com as teorias atuais. Essa classificação pode ser aceita por séculos até que algum novo conhecimento leve a uma compreensão de que as entidades compartilham outras propriedades mais fundamentalmente importantes com um tipo diferente, não com seu tipo aparente.

Às vezes, é descoberto que embora as propriedades fundamentais das entidades não sejam aquelas de seu tipo aparente,

elas também não parecem ser de qualquer outro tipo familiar. Em tais casos, uma nova estrutura teórica deve ser desenvolvida para fornecer um sistema significativo de classificação.

Existem muitas questões profundas sobre quando uma descoberta terá um grande impacto em um esquema de classificação, mas certamente um fator importante é se essa descoberta é feita no contexto de uma teoria causal coerente em que as propriedades descobertas não são apenas significativas, mas centrais (Keil, 1989, p. 159).

Essa longa citação é importante, acredito, porque demonstra a realidade mais profunda dos tipos e conceitos que a ciência descobre. Tem implicações importantes para a metodologia de OC. Em primeiro lugar, ela desafia muitas abordagens orientadas ao usuário e empiristas.

Na medida em que essa visão é correta, as relações entre dois conceitos são, portanto, relativas aos sistemas teóricos (ou paradigmas) nos quais estão inseridos. Não podemos dizer objetivamente e de uma vez por todas, por exemplo, que “esquizofrenia” e “transtornos maníaco-depressivos” são termos mais restritos (*narrower terms* - NT) em relação a “psicose”. Foi sugerido por Kringlen (1994) que a esquizofrenia etiológica pode não ser um conceito útil, enquanto que provavelmente se deveria operar alternativamente com o antigo conceito alemão “*Einheitspsychose*” (psicose unitária). Segue-se que a esquizofrenia não é NT em relação à psicose. Na medida em que isso é correto, as relações semânticas entre “esquizofrenia”, “transtornos maníaco-depressivos” e “psicose” são relações sinônimas, não relações genéricas.

Podemos concluir que as unidades básicas em OC, as relações semânticas entre dois conceitos, podem ser relativas à perspectiva e à teoria a partir da qual são considerados. Devido a esse fato, a OC não pode ser feita apenas a partir de combinações sucessivas de elementos, mas deve refletir perspectivas e teorias mais amplas.

9 Critérios de parença em OC.

Um princípio fundamental em OC é que coisas semelhantes devem ser reunidas, enquanto coisas diferentes devem ser separadas. “Pareença” é um conceito que também pode ser expresso por outros termos como “similaridade”, “uniformidade” (usado por James, 1890), “semelhança” ou “equivalência”. Muitos escritores em LIS definiram classificação e OC usando os conceitos de semelhança. Por exemplo:

- Ernest Cushing Richardson (1964, p. 1) definiu classificação como o “agrupamento de coisas semelhantes, ou, mais completamente descrito, é a organização das coisas de acordo com a semelhança e dessemelhança. Também pode ser descrito como a classificação e agrupamento de coisas”.
- Henry E. Bliss (1935, p. 3) escreveu: “Ao lidar com a multiplicidade de coisas particulares, realidades e tipos específicos, descobrimos que alguns são semelhantes, em caracteres gerais e em características específicas; e podemos conseqüentemente relacioná-los em uma classe, ou classes, ou seja, classificá-los”
- Frederick Wilfrid Lancaster (nascido em 1933) definiu a classificação como “classificar itens em 'classes conceituais'” e “formar classes de objetos com base em seu assunto”. (Lancaster, 1998, p. 17).

— Arlene G. Taylor: “A colocação dos assuntos em categorias; na organização de informações, classificação é o processo de determinar onde um pacote de informações se encaixa em uma determinada hierarquia e, em seguida, atribuir a notação associada ao nível apropriado da hierarquia para o pacote de informações e seu substituto.” (Taylor, 1999, p. 237).

Como podemos decidir a que as coisas são semelhantes? Como desenvolvemos nossos critérios de semelhança? A literatura sobre OC em LIS parece ignorar esse problema central. É porque o problema é visto como óbvio? Isso é baseado em um tipo de realismo ingênuo: que as coisas são o que parecem e que o senso imediato de semelhança das pessoas é adequado como base para a OC?

Mas quando duas coisas são iguais? O realismo ingênuo confunde aparente semelhança com semelhança em um sentido objetivo. Alguns metais podem parecer ouro, mas podem não ser metais preciosos. A análise química (não o senso comum) deve fazer uma distinção entre “o que parece ouro” e “o que é ouro”. A classificação feita para crianças pode ser baseada em atributos mais superficiais - por exemplo, "o grande livro dos trens" que considera a maioria dos aspectos do "sistema ferroviário". As classificações científicas, por outro lado, refletem algumas propriedades mais profundas, como a classificação de substâncias químicas em compostos orgânicos e inorgânicos, metais preciosos, etc., com base na teoria atômica. Quando os princípios científicos de classificação são aplicados, objetos aparentemente relacionados podem ser separados e objetos aparentemente diferentes podem ser agrupados. Por exemplo, as baleias já foram classificadas como peixes, mas hoje - influenciadas pela teoria da evolução - são classificadas como mamíferos.¹⁰ Esses exemplos demonstram que o realismo ingênuo não é adequado como método de classificação de documentos.

Se considerarmos diferentes exemplos, veremos que as relações conceituais têm diferentes tipos de motivações.

- “Instituições para Ciência da Informação” são genericamente subordinadas a “instituições”. Essa relação parece ser motivada por relações puramente lógicas (ou por relações inerentes a uma dada língua).
- “Copenhague” faz parte da “Dinamarca” (enquanto Malmö na Suécia não. Este último exemplo é motivado pelo fato de que a Dinamarca perdeu essa parte do território para a Suécia em 1658). As partes de um país são, portanto, definidas por arranjos sociais. Em outras palavras, esse tipo de relações semânticas é baseado em convenções humanas.
- As baleias são hoje classificadas como mamíferos. A explicação desta relação semântica (genérica) se deve à teoria da evolução.¹¹
- A psicologia (e também a psicopatologia) pode ser classificada como parte da neurociência (ciências naturais), como parte das ciências sociais ou como parte das humanidades (ou não). Essas diferenças são, por exemplo, visíveis na forma como esses campos são colocados nas estruturas organizacionais nas universidades. Essas classificações (e relações semânticas) frequentemente envolvem interesses profissionais. É, por exemplo, em parte uma questão de poder profissional se um campo (por exemplo, psicofarmacologia) é monopolizado por uma profissão. No nível mais profundo, a questão de saber se a psicologia é uma ciência humana ou uma ciência natural é uma questão científica relacionada a questões teóricas dentro da psicologia. É bem sabido que a psicologia está dividida sobre esta questão. Diferentes paradigmas em psicologia têm diferentes respostas. Para o

behaviorismo, a psicologia é claramente uma parte das ciências naturais (cf. Watson, 1913). Para a psicologia humanística, é claramente uma parte das humanidades. Ao classificar a psicologia como pertencente à ciência, ou ciências sociais ou humanas, estamos na verdade envolvidos em uma batalha teórica entre paradigmas. (O que a maioria das pessoas acha bastante desconfortável).

- Classificações e relações semânticas podem ser estabelecidas por generalizações empíricas. Um exemplo é Berlin e Kay (1969), que estabeleceram a generalização empírica de que as linguagens humanas refletem a classificação das sensações de cor essencialmente da mesma maneira universalmente, independentemente das diferenças históricas e culturais. Essas generalizações empíricas evitam as incômodas questões teóricas abordadas acima. Existem, entretanto, outros tipos de problemas com essa abordagem, que estão intimamente ligados às suposições básicas do empirismo.
- As classificações e relações semânticas também podem ser puramente acidentais ou ad hoc. Essas classificações podem servir a alguns propósitos muito bem. Em geral, entretanto, as classificações que refletem características essenciais nos objetos são as mais valiosas. (Este “essencialismo” não deve ser confundido com um objetivismo que ignora a atividade humana).

Podemos concluir que existem muitos tipos diferentes de critérios de parença. Eles podem ser convencionais, lógicos, psicológicos e assim por diante. Com relação aos tipos naturais, entretanto, eles devem ser vistos especialmente como critérios específicos de domínio que são descobertos pela ciência. Eles não são apenas algo que pode ser extraído de usuários ou de investigações estatísticas.

10 Métodos de Organização do Conhecimento

Por métodos de OC, entendemos os métodos de construção de sistemas de OC, como classificações e tesouros. Também os processos de indexação e classificação também são considerados métodos de OC.

Quais são os métodos de classificação e OC? Podemos primeiro fazer uma distinção entre a classificação nas ciências (como biologia ou arqueologia) e em LIS. Este artigo é sobre OC em LIS. No entanto, devemos argumentar que basicamente os métodos de OC em LIS estão ligados aos mesmos paradigmas fundamentais em epistemologia que os métodos de classificação em ciências e outros campos. Primeiro, no entanto, vamos considerar brevemente alguns métodos:

- Estandarização/uniformização
- OC baseado em computador
- Métodos “manuais” ou “intelectuais”
- Métodos quantitativos
- Métodos qualitativos
- Métodos baseados em texto
- Métodos baseados em pessoas
- Métodos baseados em instituições (por exemplo, estudos de organizações universitárias)
- Métodos bibliométricos
- Métodos baseados em frequência de palavras

- Métodos sociológicos
- Métodos históricos
- Métodos pragmáticos, epistemológicos e críticos

É importante perceber que uma qualidade importante de um sistema de OC pode ser sua função como padrão. Frequentemente, as classificações e outros sistemas são feitos como compromissos entre especialistas em um campo e declarados como um padrão. Argumentos científicos entre pesquisadores da área não são publicados e discutidos na literatura como parte desse processo. Este método é epistemologicamente uma espécie de “autoritarismo”: a qualidade do sistema é reivindicada por referência à sua autoridade como padrão. No entanto, como pesquisadores em OC, temos que considerar outras qualidades de OC também e não devemos aceitar o “autoritarismo” como um critério de conhecimento científico.

Uma distinção, que pode ser encontrada na literatura, é entre métodos baseados em computador e métodos “manuais” de OC. O termo “manual” é um pouco peculiar, porque o que significa são métodos mais intelectuais. Essa distinção não é em si uma distinção fundamental. Se for possível desenvolver algumas regras explícitas sobre como classificar documentos, essas regras podem ser usadas por humanos ou por computadores. Em outras palavras, às vezes os humanos não têm um desempenho melhor do que os computadores, que seguem instruções bastante simples. Em outras ocasiões, no entanto, os humanos aplicam um conhecimento profundo, que não está disponível para os computadores. E às vezes os computadores processam grandes quantidades de dados que não são possíveis para os humanos processar. Portanto, não devemos considerar métodos baseados em computador e métodos manuais como uma distinção metodológica básica. Devemos antes classificar os métodos de OC de outras maneiras, perguntando que tipo de regras, que tipo de conhecimento e que tipo de processos cognitivos estão envolvidos no processo? Em geral, nosso conhecimento de como os humanos classificam é limitado.¹²

Outra distinção é entre métodos quantitativos e qualitativos. Os métodos quantitativos podem, por exemplo, ser baseados em frequências de palavras. Os métodos qualitativos podem, por outro lado, basear-se em interpretações de significado. Novamente, não considero essa uma distinção fundamental. É sabido que qualquer investigação quantitativa pressupõe uma análise qualitativa adequada. Os métodos qualitativos, por outro lado, podem resultar em uma série de categorias e, portanto, aproximar-se da abordagem quantitativa. O sociólogo Pierre Bourdieu usa muitas estatísticas (dados quantitativos) em sua pesquisa; no entanto, ele nega explicitamente fazer isso de uma forma positivista. Existem, portanto, maneiras importantes de usar dados que vão mais fundo do que a distinção entre abordagens quantitativas e qualitativas.

Os sistemas de OC podem usar dados da literatura, de pessoas, de organizações, etc., de muitas maneiras diferentes e em muitas combinações. As abordagens bibliométricas baseiam-se principalmente em literaturas, assim como muitos métodos convencionais e “manuais” (incluindo o conhecido princípio de identificação da *garantia literária* para uma determinada classe ou descritor). O uso de conhecimento histórico sobre o desenvolvimento de campos de conhecimento (conforme exemplificado em Wallerstein, 1996 e Hjørland, 2000c) é uma forma diferente de usar a literatura como dados para OC. Esta última abordagem também usa o desenvolvimento da organização das universidades como base para a construção de OC.

Pessoas também podem ser usadas como fontes nas quais basear o desenvolvimento de OC. Em ciência cognitiva e inteligência artificial, os métodos de *elicitação de conhecimento* são bem conhecidos (por exemplo, Cooke, 1994). Este é, em minha opinião, um *método indireto* porque temos que conhecer os métodos pelos quais os especialistas, de quem

obtemos o conhecimento, o obtiveram em primeiro lugar. Em outras palavras, devemos ser capazes de argumentar quais afirmações de conhecimento são mais bem fundamentadas. Se não resolvermos esse problema, estaremos apenas indiretamente qualificados para construir e avaliar sistemas de OC. No entanto, as pessoas também podem ser usadas como fontes de outras maneiras. Elas podem, por exemplo, ser vistas como membros de *comunidades de discurso* e ser estudadas como tais. Podemos estudar a divisão social do trabalho e a dependência entre diferentes pessoas e grupos de pessoas (compare Whitley, 1984).

Quando vamos decidir em que tipo de fontes construir (literatura, pessoas, organizações), precisamos saber algo sobre como o conhecimento relevante é distribuído. Os métodos de elicitación de conhecimento parecem ser construídos na suposição de que o conhecimento necessário está disponível dentro de um grupo de especialistas ou outras pessoas. Isso pode ser mais ou menos o caso. No entanto, abordagens mais acadêmicas baseiam-se em outras suposições. O significado das palavras é rastreado em grande detalhe e registrado em dicionários históricos com citações e citações de autores autorizados. Nesse caso, as informações relevantes não devem estar prontas à mão na mente de especialistas, mas algo que pode ser construído trabalhando com a literatura (não apenas como uma espécie de pesquisa de opinião). A questão da autoridade cognitiva é obviamente extremamente importante e difícil. O trabalho acadêmico não é apenas “subjetivo”, escolhendo as autoridades de acordo com as preferências pessoais. Quanto mais qualificado, menos “subjetivo” neste sentido da palavra. A “objetividade” da decisão do trabalho acadêmico é parcialmente baseada no conhecimento de diferentes tradições e metaperspectivas e na evidência histórica acumulada. Em que poderia talvez dizer que uma diferença fundamental entre empirismo / racionalismo e historicismo / pragmatismo é a ênfase na leitura no processo de pesquisa.

Basicamente, os métodos de OC estão relacionados às teorias fundamentais da epistemologia. Todos os pesquisadores de qualquer área são sempre mais ou menos influenciados por certos ideais de como obter conhecimento. Este também é o caso em relação à construção de sistemas de OC. Se os alunos da LIS fizeram um curso de filosofia da ciência, eles devem ser capazes de identificar a abordagem epistêmica dominante em qualquer artigo sobre OC (e também em qualquer sistema de OC). Alguns artigos e sistemas são baseados principalmente em generalizações empíricas. Sistemas baseados em medidas de frequência de palavras podem ser o melhor exemplo. Outros papéis e sistemas são baseados principalmente em regras e deduções racionais (embora muitas vezes ignorem questões empíricas). Os sistemas analíticos facetados na tradição de Ranganathan podem fornecer os melhores exemplos. Um terceiro tipo de sistema é baseado no estudo da evolução dos campos do conhecimento. Até certo ponto, pode-se dizer que os sistemas baseados em disciplinas como Dewey (CDD) estão em conformidade com o ideal. Por outro lado, a atualização desse sistema é cada vez mais influenciada pelos princípios da tradição analítico-facetária (conforme discutido por Miksa, 1998). E, em certo sentido, a CDD também é empiricamente baseada na literatura, que é classificada pelo sistema. A CDD não é, portanto, tão clara em seus princípios metodológicos como os exemplos anteriores. O quarto ideal básico é a epistemologia pragmática. Neste caso, os sistemas são desenvolvidos com base na análise de objetivos, valores e consequências. Por exemplo, as pesquisadoras podem considerar o objetivo da emancipação feminina como o objetivo principal da bolsa de estudos feminista. Elas podem, portanto, considerar importantes quaisquer relações que apoiem esse objetivo. “Similaridade” é assim identificada por classes de documentos que compartilham funções semelhantes em relação a este objetivo. A classificação pragmática também pode ser chamada de classificação crítica ou política. Pode parecer à primeira vista estranho e oposto às normas científicas comuns.

Epistemologia pragmática e OC pragmática não significam que uma pessoa (ou um campo inteiro) pode simplesmente fazer as coisas da maneira que se adapte aos seus interesses pessoais (ou aos interesses dos pesquisadores no campo). Se isso for feito, se a pesquisa apenas produzir “construções sociais”, a realidade tornará essas construções incoerentes. Eles serão contestados por argumentos empíricos e teóricos. A produção de “conhecimento” incoerente não tem valor e não pode ser um objetivo sério. Portanto, a filosofia pragmática está ligada a uma forma de realismo. O método pragmático não se opõe a aspectos do empirismo, racionalismo e historicismo. Afirma, no entanto, que nenhuma evidência isolada é suficiente. Os critérios finais da verdade estão ligados aos objetivos e atividades humanas. Você não pode evitar considerar essas questões, embora possam parecer desconfortáveis.

O fundador do pragmatismo, Charles Sanders Peirce (1839-1914), escreveu:

O significado racional de cada proposição está no futuro. Como assim? O significado de uma proposição [seu interpretante lógico] é em si uma proposição. Na verdade, não é outra coisa senão a própria proposição de que é o significado: é uma tradução dele. Mas, das miríades de formas nas quais uma proposição pode ser traduzida, qual é aquela que deve ser chamada de seu próprio significado? É, de acordo com o pragmaticista, aquela forma pela qual a proposição se torna aplicável à conduta humana, ... aquela forma que é mais diretamente aplicável ao autocontrole em todas as situações e para todos os fins. É por isso que ele localiza o significado no tempo futuro; pois a conduta futura é a única que está sujeita ao autocontrole (Peirce, 1905).

Desse modo, a epistemologia pragmática vê o significado das palavras como conectado aos atos de fala e aos objetivos que os humanos tentam satisfazer por meio de seus atos. Da mesma forma, considera as relações e classificações semânticas determinadas por suas funções como ferramentas para os objetivos humanos. Como afirmou o filósofo pragmático John Dewey:

Até agora, o nominalismo e o conceitualismo - a teoria de que os tipos existem apenas em palavras ou em ideias - estavam no caminho certo. Enfatizou o caráter teleológico dos sistemas e classificações, que existem por uma questão de economia e eficiência no alcance dos fins. Mas essa verdade foi pervertida em uma noção falsa, porque o lado ativo e de fazer da experiência foi negado ou ignorado.

As coisas concretas têm *modos* [ênfase no original] de agir, tantos modos de agir quantos pontos de interação com outras coisas. Uma coisa é insensível, indiferente, inerte na presença de outras coisas; está alerta, ansioso e agressivo em relação a outras coisas; em um terceiro caso, é receptivo, dócil. Agora, diferentes maneiras de se comportar, apesar de sua diversidade infinita, podem ser classificadas em conjunto em vista de uma relação comum para um fim. Nenhuma pessoa sensata tenta fazer tudo. Ele tem certos interesses principais e objetivos principais pelos quais torna seu comportamento coerente e eficaz. Ter um objetivo é limitar, selecionar, concentrar, agrupar. Assim, uma base é fornecida para selecionar e organizar as coisas de acordo com suas maneiras de agir, relacionadas com o prosseguimento da busca. As cerejeiras serão agrupadas de forma diferente por marceneiros, pomares, artistas, cientistas e foliões. Para a execução de diferentes propósitos, diferentes formas de agir e reagir por parte das árvores

são importantes. Cada classificação pode ser igualmente válida quando a diferença dos fins é levada em consideração.

No entanto, existe um padrão objetivo genuíno para a bondade das classificações especiais. Um ajudará o marceneiro a chegar ao seu fim, enquanto outro o atrapalhará. Uma classificação ajudará o botânico a levar avante seu trabalho de investigação, e outra o retardará e confundirá. A teoria teleológica da classificação não nos compromete, portanto, com a noção de que as classes são puramente verbais ou puramente mentais. A organização não é mais meramente nominal ou mental em qualquer arte, incluindo a arte da investigação, do que em uma loja de departamentos ou sistema ferroviário. A necessidade de execução fornece critérios objetivos. As coisas devem ser classificadas e organizadas de modo que seu agrupamento promova uma ação bem-sucedida para os fins. Conveniência, economia e eficiência são as bases da classificação, mas essas coisas não se restringem à comunicação verbal com os outros nem à consciência interior; dizem respeito à ação objetiva. Eles devem ter efeito no mundo.

Ao mesmo tempo, uma classificação não é uma simples transcrição ou duplicata de algo acabado e feito para arranjos pré-existentes na natureza. É antes um repertório de armas para atacar o futuro e o desconhecido. Para o sucesso, os detalhes do conhecimento passado devem ser reduzidos de fatos nus a significados, quanto menos, mais simples e mais extensos, melhor ... (Dewey, 1948, p. 151-154).

A Figura 5 abaixo resume os quatro métodos fundamentais de OC para os quais argumentei. É importante perceber que são idealizações. Eles não existem e não podem existir em formas puras. Qualquer procedimento empírico deve envolver tipos de evidências lógicas e não empíricas. Todas as regras devem ser aplicadas a alguma realidade empírica. Todos os métodos empíricos e racionais devem ser baseados em pré-entendimentos e significados transmitidos através da linguagem e produtos culturais. E qualquer tipo de pragmatismo é limitado por restrições estabelecidas pelo mundo real por meio de evidências empíricas.

Ao conhecer os principais pontos fortes e fracos das posições epistemológicas básicas, o especialista em informação está equipado com algum conhecimento geral dos lados fortes e fracos dos diferentes métodos de OC.

	<i>“Classificação Científica”</i>	<i>“Classificação Bibliográfica”</i>
Empirismo (Observações e induções)	Classificação fornecida por generalizações estatísticas (por exemplo, análise fatorial) com base em "similaridade". Exemplos: Classificações de doenças mentais em psiquiatria (DSMIV) tipos de inteligência em psicologia com base em análises estatísticas de resultados de testes.	Documentos agrupados com base em algum tipo de semelhança, por exemplo, termos comuns em RI tradicional ou acoplamento bibliográfico. Exemplos: “Atlas da ciência” e visualizações (White & McCain, 1998). “Frentes de Pesquisa” I SCI e algoritmos para recuperação da informação.
Racionalismo (Princípios da razão pura. Deduções)	Classificação baseada em divisões lógicas universais. Exemplos: Sistemas baseados em Frame em Inteligência Artificial. A análise de Chomsky da estrutura	Análise de facetas baseada em divisões lógicas e "categorias eternas e imutáveis" Exemplos: Ranganathan, Bliss II e Langridge. Redes semânticas. De

	profunda da linguagem	acordo com Miksa (1996), a CDD tem usado cada vez mais esta abordagem.
Historicismo (Estudo de contexto e desenvolvimento - Explicando pré-compreensão)	Classificação baseada no desenvolvimento histórico ou evolutivo. Exemplos: Taxonomias biológicas baseadas na teoria da evolução. Classificação das ciências com base na sua história e estruturas organizacionais.	Sistemas baseados no estudo do desenvolvimento do conhecimento e comunidades produtoras de conhecimento (a divisão social do trabalho (científico)). Exemplos: Wallerstein (1996) (ver também Hjørland, 2000c). A característica da CDD é que distribui assuntos por disciplina.
Pragmatismo (Análise de metas, valores e consequências tanto no sujeito quanto no objeto)	Classificações baseadas em valores, políticas e objetivos específicos, por ex. epistemologia feminista.	Sistemas baseados em “garantia cultural” ou “classificação crítica”. Exemplos: os enciclopedistas franceses, os marxistas, classificações que servem a coleções feministas.

Figura 5: Métodos Fundamentais de Classificação

11 Conclusão

Este artigo descreveu os aspectos fundamentais da OC como um campo de estudo. As suposições e atitudes básicas são que as unidades básicas de OC são relações semânticas entre conceitos. Essas relações semânticas não podem ser estabelecidas primariamente por suposições universalistas, mas principalmente devem ser entendidas como específicas do domínio, conforme descobertas por (e construídas por) disciplinas científicas. OC em LIS não pode ignorar conceitos, teorias e descobertas em disciplinas específicas, e os métodos de OC em LIS estão no nível mais profundo com base nas mesmas suposições filosóficas que os métodos da ciência e da erudição. Isso implica que a discussão fundamental da base da OC em LIS está fortemente conectada à discussão de diferentes teorias da epistemologia. Estudos epistemológicos têm sido raros em LIS, e parece urgente para nosso campo atualizar nossas qualificações nesta área.

Notas

- 1 Este trabalho é uma versão ligeiramente modificada de um artigo apresentado em Salamanca, Espanha (Hjørland, 2003a).
- 2 A partir de 1937 denominado *Fédération Internationale de Documentation*, FID, e a partir de 1986 *International Fédèration for Information and Documentation*, FID.
- 3 A história do periódico de resumo remonta, entretanto, a 1665, cf. Manzer (1977).
- 4 A ciência da computação trata de “tecnologia da informação”, TI ou processamento de dados. A ciência da computação é chamada na Dinamarca de “datalogi”, a ciência dos dados.
- 5 Escrevo isto ciente de que o filósofo pragmático John Dewey usou a metáfora do espelho: “Uma classificação de livros para ser eficaz no lado prático deve corresponder às relações de assuntos, e essa correspondência pode ser assegurada apenas quando a organização intelectual ou conceitual é baseada na ordem inerente

- aos campos do conhecimento, que por sua vez reflete a ordem da natureza. ” (Dewey, 1929, p. Viii)
- 6 Considerando que a semiótica tradicional tende a ser formalista e a abstrair os signos de seus contextos de uso; A semiótica social examina as práticas semióticas, específicas para uma cultura e comunidade, para a construção de vários tipos de textos e significados em vários contextos situacionais e contextos de atividade culturalmente significativa. Para uma introdução à semiótica social, ver, por exemplo, Hodge e Kress (1988). As formulações neste artigo foram inspiradas por Karpatschhof (2000).
 - 7 Às vezes, é feita uma diferenciação entre a organização do conhecimento e a organização de outras “coisas” (incluindo objetos físicos, processos, fenômenos como a dor etc.). Outras coisas além do conhecimento (reivindicações) não se classificam, no entanto. Eles também são classificados por nossa conceituação de mundo.
 - 8 Na literatura da ciência cognitiva (por exemplo, Keil, 1989), é feita uma distinção entre significados e conceitos de palavras, ou entre estrutura de conceito e estrutura semântica. As pessoas são descritas como possuidoras de dois sistemas diferentes: um sistema semântico de significados de palavras e um sistema de conceitos. Esta distinção não é feita no presente artigo.
 - 9 “A semântica estruturalista saussuriana acredita que a melhor maneira de descrever o significado é descrever a delimitação mútua de conceitos (relações semânticas). A Terminologia Tradicional acredita que a melhor maneira de descrever conceitos é determinar sua posição em um sistema de conceitos que visualiza relações lógicas e ontológicas. Com base na posição no sistema de conceitos, uma definição será formulada ”(Temmermann, 1997, p. 53).
 - 10 Aristóteles percebeu que golfinhos e botos tinham pulmões. Eles não botavam ovos e alimentavam seus filhotes por nascer por meio de uma placenta. Ele os agrupou com os quadrúpedes em vez dos peixes. (Aristóteles era uma voz solitária. Por mais de 2.000 anos depois, sua ideia se perdeu, e as baleias eram peixes. Somente em tempos relativamente modernos as baleias foram colocadas de volta com os mamíferos.)
 - 11 Reveal (1999) descreve diferentes paradigmas, métodos e ferramentas usados na taxonomia de plantas e menciona, entre outras coisas, o uso de (1) anatomia: 1800s, (2) números de cromossomos: 1920's, (3) estudos comparativos de jardins: 1940's, (4) palinologia: 1950, (5) fenética: 1960, (6) sistemática bioquímica: 1970, (7) cladística: 1980 e (8) DNA e RNA do cloroplasto: 1990.
 - 12 Klumpner (1993, p. 1) relata: “Esses problemas levaram o Indexing Study Group da American Psychoanalytic Association a um experimento. Cerca de uma dúzia de analistas experientes indexaram independentemente uma passagem da *Standard Edition* [as obras de Sigmund Freud]. Quando compararam o que haviam feito, todos concordaram que a falha em concordar sobre quais termos indexar era humilhante e impressionante. O grupo não chegou a acordo sobre quais palavras obrigatórias *ver* ou *ver também* as diretrizes, ou sobre as palavras que deveriam seguir essas diretrizes. Um beco sem saída? ”. O próprio Klumpner tentou resolver o problema por meio de um estudo quantitativo dos índices efetivamente produzidos. O que eu gostaria de enfatizar é que, para progredir, precisamos desse tipo de desacordo registrado e discutido qualitativamente na literatura. Eles são extremamente raros, razão pela qual afirmo que sabemos pouco sobre o OC “manual”.

REFERENCES

- Albrechtsen, H.; Andersen, H.H.K.; Bødker, S.; Pejtersen, A.M.,(2001). *Affordances in Activity Theory And Cognitive Systems Engineering*. Denmark. Risø: Available at: <http://www.risoe.dk/rispubl/SYS/syspdf/tris-r-1287.pdf>
- Andersen, H. (1999). Political Attitudes and Cognitive Convictions among Danish Social Science Researchers. *Scientometrics*, 46(1), 87-108.
- Arbib, M. A.; Kfoury, A.J. & Moll, R. N. (1981). *Basis For Theoretical Computer Science*. Berlin: Springer-Verlag.
- Beaugrande,R. de (1997). On history and historicity in modern linguistics. Formalism versus functionalism revisited. *Functions of Language*, 4/2, 169-213. <http://www.beaugrande.com/>
- Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- Bernal, J. D. (1948). Preliminary Analysis of Pilot Questionnaire on the Use of Scientific Literature. In: Royal Society [Great Britain]. The Royal Society Scientific Information Conference: Report and Papers Submitted. 1948 21 June – 2 July; London, England. London, England: Royal Society. Pp. 589-637.
- Bliss, H. E. (1935). *A System of Bibliographical Classification*. New York: The H. W. Wilson Company.
- Borko, H. (1968). Information science: What is it? *American Documentation*, 19(1), 35.
- Bradford, S. C. (1950). *Documentation*. London: C. Lockwood.
- Buckley, G. (2001). *Semantics*. http://www.ling.upenn.edu/courses/Spring_2001/ling001/semantics.html [visited 2003-01-08]
- Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human- Computer Studies*, 41(6) , 801-849.
- Dahlberg, I. (1991). Philosophical foundations of conceptual ordering systems. *Advances in Knowledge Organization*, 3, 102-119.
- Daly, L. W. (1967). *Contributions to a History of Alphabetization in Antiquity and the Middle Ages*, Brussels: Latomus.
- Dewey, J. (1929). Introduction. IN: H. E. Bliss: *The Organization Of Knowledge And The System Of The Sciences*. New York, Holt (pp. vii-ix).
- Dewey, J. (1948). *Reconstruction in Philosophy*. Enlarged edition. New York: Beacon. (Original work published 1920)
- Dewey, M. (1979): *Dewey Decimal Classification and Relative Index*. Edition 19. Albany, N.Y.: Forest Press, 1979. Vol. 1.
- Dogan, M. (2001). Paradigms in the Social Sciences. IN: *International Encyclopedia of the Social & Behavioral Sciences*, (vol. 16, pp. 11023-11027). Ed. By N. J. Smeltser & P. B. Baltes. Amsterdam: Elsevier.
- Ellis, B. (1990). *Truth and Objectivity*. Oxford: Basil Blackwell.
- Fjordback Søndergaard, T.; Andersen, J. & Hjørland, B. (2003). Documents and the Communication of Scientific and Scholarly Information. Revising and Updating the UNISIST Model. *Journal of Documentation*, 59(3), 278-320.
- Frohmann, B. (1994). The Social Construction of Knowledge Organization: The Case of Melvin Dewey. *Advances in Knowledge Organization*. 4, 109-117.
- Griffith, B. C. (Ed.). (1980). *Key Papers in Information Science*. New York: Knowledge Industry Publications.

- Hargens, L. L. (2000). Using the literature: reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, 65 (December), 846-865.
- Hjelmslev, L. (1943) *Omkring sprogteoriens grundlæggelse*. København: B. Lunos bogtrykkeri a/s. (Many later editions and translations, e.g. *Prolegomena to a Theory Of Language*. translated by Francis J. Whitfield. Baltimore : Waverly Press, 1953).
- Hjørland, B. (1988). Information Retrieval in Psychology. *Behavioral and Social Sciences Librarian*, Vol 6 (3/4), 1988, 39-64.
- Hjørland, B. (1992). The concept of “subject” in Information Science. *Journal of Documentation*, vol. 48, no. 2, 172-200.
- Hjørland, B. (1993). *Emnerepræsentation og informationssøgning. Bidrag til en teori på kundskabsteoretisk grundlag*. Göteborg: Valfrid,
- Hjørland, B. (1997): *Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science*. Westport & London: Greenwood Press.
- Hjørland, B. (1998a). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25(1/2), 16-31.
- Hjørland, B. (1998b). The Classification of Psychology: A Case Study in the Classification of a Knowledge Field. *Knowledge Organization*, 24(4), 162-201.
- Hjørland, B. (2000a). Documents, Memory Institutions, and Information Science. *Journal of Documentation*, vol. 56(1), 27-41
- Hjørland, B. (2000b). Library and Information Science: Practice, theory, and philosophical basis. *Information Processing and Management*, 36(3), 501-531.
- Hjørland, B. (2000c). Review of Wallerstein et al. (1996). Open the Social Sciences, report of the Gulbenkian Commission on the Restructuring of the Social Sciences. Stanford, Calif.: Stanford University Press. I: *Knowledge Organization*, 27(4), 238-241
- Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology*. 52(9):774–778.
- Hjørland, B. (2002a). Domain analysis in information science. Eleven approaches - traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462.
- Hjørland, B. (2002b), Epistemology and the Socio-Cognitive Perspective in Information Science. *Journal of the American Society for Information Science and Technology*, 53(4), 257-270.
- Hjørland, B. (2002c) The Methodology of Constructing Classification Schemes. A Discussion of the State-of-the-Art. IN: *Proceedings of the Seventh International ISKO Conference 10.-13. July 2002, Granada, Spain*. Ed. By María J. López-Huertas. Würzburg, Germany: Ergon Verlag, pp. 450-456. (Advances in Knowledge Organization, Vol. 8).
- Hjørland, B. (2002d). Principia Informatica. Foundational Theory of Information and Principles of Information Services. IN: *Emerging Frameworks and Methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)*. Ed. By Harry Bruce, Raya Fidel, Peter Ingwersen, and Pertti Vakkari. Greenwood Village, Colorado, USA: Libraries Unlimited. (Pp. 109-121).
- Hjørland, B. (2003a). Fundamentals of knowledge organization. IN: *Tendencias de investigación en organización del conocimiento / Trends in Knowledge Organization Research*. (Pp. 83-116; +English abstract p. 51 + Spanish abstract p. 19). José Antonio Frias & Crispulo Travieso (Eds.). Salamanca, Spain: Ediciones Universidad de Salamanca. Presentation available at: <http://www.db.dk/bh/publikationer/Filer/Presentation.ppt>
- Hjørland, B. (2003b), Social And Cultural Awareness and Responsibility in Library, Information and Documentation Studies. IN: *Aware and Responsible: Papers of the Nordic-International Colloquium on Social And Cultural Awareness and Responsibility in*

- Library, Information and Documentation Studies (SCARLID)*. Edited by W. Boyd Rayward, Joacim Hansson, Vesa Suominen. Lanham, MD: Scarecrow Press, (in press).
- Hjørland, B. & Albrechtsen, H. (1995). Toward A New Horizon in Information Science: Domain Analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Hjørland, B. & Kylesbech Nielsen, L. (2001). Subject Access Points in Electronic Retrieval. *Annual Review of Information Science and technology*, vol. 35, 249-298.
- Hjørland, B. & Sejer Christensen, F. (2002). Work tasks and socio-cognitive relevance. A specific Example. *Journal of the American Society for Information Science and Technology*, 53(11), 960-965.
- Hodge, R. & Kress, G. (1988). *Social Semiotics*. New York; Cornell University Press.
- James, W. (1890). *The Principles of Psychology*. (2 vols.). New York: Henry Holt.
- Karpatschof, B. (2000). *Human activity. Contributions to the Anthropological Sciences from a Perspective of Activity Theory*. Copenhagen: Dansk Psykologisk Forlag. ISBN: 87-7706-311-2
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, Massachusetts: The MIT Press.
- Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation* vol. 16, no. 3, 223-233.
- Klumpner, G. H. (1993). *A Guide To The Language Of Psychoanalysis. An Empirical Study Of The Relationships Among Psychoanalytic Terms And Concepts*. Madison, Connecticut: International University Press, Inc.
- Kringlen, E. (1994). Is the concept of schizophrenia useful from an aetiological point of view? A selective review of findings and paradoxes. *Acta Psychiatrica Scandinavica*, 90 (suppl. 384), 17-25.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press. 2nd. Eition 1970. 3rd edition 1996.
- Lancaster, F. W. (1998). *Indexing And Abstracting In Theory And Practice*. 2nd ed. London: Library Association.
- Manzer, B. M. (1977). *The Abstract Journal, 1790-1920: Origin, Development and Diffusion*. Metuchen, N.J.: The Scarecrow Press.
- Miksa, F. L. (1998). *The DDC, The Universe of Knowledge, and the Post-Modern Library*. Albany, NY: Forest Press.
- Pao, M. L. (1993). Term and citation retrieval: A field study. *Information Processing & Management*, 29(1), 95112.
- Pao, M. L., & Worthen, D. B. (1989). Retrieval effectiveness by semantic and pragmatic relevance. *Journal of the American Society for Information Science*, 40(4), 226235.
- Peirce, C. S. (1905). What pragmatism is. *The Monist*, 15, 161-181.
- Poulsen, C. (1994). *Informationens skygge og foran: informationskvalitet, informationsekspllosion og online kataloger*. Roskilde, Denmark: Institut for Datalogi, Kommunikation og Uddannelsesforskning, Roskilde Universitetscenter.
- Rees-Potter, L. K. (1987). *A bibliometric analysis of terminological and conceptual change in sociology and economics with the application to the design of dynamic thesaural systems. Vol. 1-2*. Ph.D dissertation. *Dissertation Abstracts International*, 48(06), SecA, p.1344.
- Rees-Potter, L. K. (1989). Dynamic thesaural systems: a bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management*, 25(6), 677- 691.

- Rees-Potter, L. (1991). Dynamic thesauri: the cognitive function. Tools for knowledge organisation and the human interface. *Proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990*.
- Part 2, 1991, 145-150. Reveal, J. L. (1999). PBIO 250 Lecture Notes. Principles of Plant Taxonomy. <http://www.inform.umd.edu/PBIO/pb250/prin.html> (hentet 10-02-2003)
- Richardson, E. C. (1964). *Classification. Theoretical and practical. Third edition*. Hamden, Connecticut: The Shoe String Press, Inc. (Reprinted unaltered from 1930 edition).
- Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.
- Schneider, J. W. & Borlund, P. (2002). Preliminary study of the potentiality of bibliometric methods for the construction of thesauri. IN: *Emerging Frameworks and Methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)*. Ed. By Harry Bruce, Raya Fidel, Peter Ingwersen, and Pertti Vakkari. Greenwood Village, Colorado, USA: Libraries Unlimited. (Pp. 151-165).
- Taylor, A. G. (1999). *The Organization of Information*. Englewood, Colorado: Libraries Unlimited, Inc.
- Temmerman, R. (1997). Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology. *Hermes. Journal of Linguistics*, 18, 51-93. Available at: http://hermes.asb.dk/archive/FreeH/H18_04.pdf (Visited 2003-01-11).
- Törnebohm, H. (1974). *Paradigm i vetenskapernas värld och i vetenskapsteorien*. Göteborg: University of Göteborg.
- Unisist (1971), *Study Report on the feasibility of a World Science Information System*. By the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. Paris, Unesco.
- U.S. Department of Labor. Bureau of Labor Statistics (2002). *Standard Occupational Classification (SOC) System*. Washington. D.C. <http://www.bls.gov/soc/>(Visited December 30, 2002).
- Wallerstein, I. (1996). *Open the Social Sciences, report of the Gulbenkian Commission on the Restructuring of the Social Sciences*. Stanford, Calif.: Stanford University Press.
- Watson, J. B. (1913) Psychology as the behaviorist views it. *Psychological Review*, 20, 158-177.
- Whitley, R.(1984). *The Intellectual and Social Organizations of the Sciences*. Oxford University Press.
- Ørom, A. (2003) The organization of knowledge about art. Paper to appear in *Knowledge Organization 2003*.