

As Ciências Documentais e a Construção da Web Semântica

Cristina Ribeiro

Resumo

Na investigação e na prática, os especialistas das Ciências Documentais têm ao longo do tempo desenvolvido abordagens à caracterização, descrição, organização e exploração dos documentos. A evolução tecnológica das últimas décadas trouxe duas vagas de modificações de fundo ao trabalho nesta área. A primeira chegou com a disponibilidade de meios poderosos de armazenamento e tratamento automático dos documentos, a segunda corresponde à vulgarização das comunicações em rede e ao aparecimento da Web. A mudança mais significativa aqui é o alargamento da produção e do uso de todo o tipo de documentos aos não especialistas. Neste cenário rapidamente se verificou que a principal limitação ao acesso aos documentos provinha da falta dos modelos e organização que caracterizavam as colecções tradicionais.

Está neste momento a surgir, na comunidade das tecnologias para a Web, uma nova visão sobre os documentos centrada no seu significado. O núcleo dos desenvolvimentos nesta área está no consórcio W3C e na sua iniciativa "Semantic Web". O objectivo das propostas que surgem nesta área é tornar a informação na Web inteligível para as máquinas, explicitando o seu significado e estabelecendo linguagens para os conceitos usados. Os resultados das iniciativas no sentido da Web semântica dependem do empenhamento dos investigadores das Ciências da Informação, que podem simultaneamente contribuir com o seu saber e experiência de casos, obter os meios para experimentar as novas soluções que serão requeridas pelo ambiente dinâmico da Web e colaborar no estabelecimento de padrões para a informação que é partilhada.

Abstract

Both in research and in practice, Information Science specialists have long been developing approaches to the characterization, description, organization and use of documents. Technological evolution in the past few decades brought about two cycles of deep change in this area. The first has come with the generalized availability of powerful tools for storing and manipulating documents, the second is associated with the rapid growth of network communications and the Web. The most significant change is the opening of the world of document production and use to the non-specialists. In this context it has been quickly noticed that the main constraint on document access came from the absence of the models and organization, typical of traditional collections.

A new vision on documents, centered on their meaning, is currently emerging in the community of Web technologies. The core developments in this area come from the W3C Consortium and its "Semantic Web" initiative. The goal of the proposals in this area is to make information on the Web "machine understandable". The success of the initiatives towards the Semantic Web depends on the commitment of experts from Information Science, who can contribute with their knowledge and experience from case studies, while obtaining the tools to experiment new solutions required by the dynamic Web environment and cooperating in setting the standards for the information to be shared.

Os Documentos nas Ciências Documentais e na Informática

As Ciências Documentais e a Informática têm sempre tido como objecto em comum a Informação, com abordagens muito diversas à sua manipulação. Nas Ciências Documentais valorizou-se a entidade documento como objecto de estudo, e o tratamento da informação aparece subordinado ao dos documentos. A ênfase no tratamento de documentos fez desenvolver métodos para os caracterizar e classificar muito dependentes da análise por especialistas. Na Informática, a abordagem centrada na tecnologia levou a valorizar a informação muito estruturada, elegendo como objectos os itens de informação elementares e os seus relacionamentos. A ênfase no tratamento automático fez centrar o estudo mais na informação do que nos documentos de onde provinha.

A generalização da divulgação de informação em linha e o uso generalizado da Web alterou a relação com a informação em ambos os domínios. Do lado das Ciências Documentais há um enorme incremento do uso da tecnologia, necessária tanto para o avanço na caracterização dos documentos como para a sua divulgação nos novos meios. Do lado da Informática começou a ser necessário alargar o âmbito dos objectos tratados para incluir informação proveniente de fontes diversas e naturalmente pouco estruturada.

A ligação da informação ao documento, tradicional nas Ciências Documentais, tem um aspecto intrínseco muito importante: a manutenção do contexto de origem do documento. Este é um aspecto de importância crescente também na informatização das organizações, em que uma parcela cada vez maior dos documentos é objecto de análise, modelização, aquisição da informação subjacente e eventual preservação.

Uma diferença tradicional entre a manipulação da informação nas Ciências Documentais e na Informática tem a ver com a fase do processo em que se faz a modelização da informação. A estruturação da informação tradicional na Informática é prévia à sua aquisição, e o modelo adoptado enforma as aplicações suportados pela informação em causa. Nas Ciências Documentais a modelização ocorre tradicionalmente após a recolha da informação e da análise dos processos que a produziram.

A aproximação que se tem observado na prática entre os profissionais das Ciências Documentais e da Informática decorre de necessidades muito concretas sentidas em ambas as áreas. A necessidade de tratar objectos mais complexos, em

novos suportes e com componente tecnológica apreciável requer aos profissionais das Ciências Documentais uma reconstrução dos seus modelos e técnicas. A crescente valorização de informação com carácter pouco estruturado, requerendo modelos mais abrangentes, bem como a importância da preservação do contexto, têm levado a Informática a debruçar-se sobre as teorias e a prática do tratamento dos documentos.

Em ambas as áreas se evoluiu no sentido de caracterizar a informação nos seus aspectos fundamentais e de clarificar a relação entre informação e documento.

As Tecnologias

A partir dos anos 60, a disponibilidade de meios de armazenamento e tratamento automático dos dados fez aparecer os primeiros modelos e sistemas de bases de dados. Os desenvolvimentos conceptuais e o aumento da sofisticação dos sistemas oferecidos não pararam desde aí, acompanhando o desenvolvimento tecnológico e o alargamento das áreas de aplicação. Hoje em dia as bases de dados suportam todos os domínios de actividade em que seja necessário a persistência de informação e uma parte substancial das aplicações informáticas incluem-nas de forma implícita ou explícita.

Os meios computacionais abriram novas possibilidades aos profissionais das Ciências Documentais. A compilação em bases de dados da informação de catalogação, as possibilidades de pesquisa em grandes volumes de informação e as ferramentas de apoio à edição ilustram a mudança qualitativa nesta área, considerando apenas o apoio às tarefas tradicionais.

A evolução no uso dos computadores trouxe também uma vaga de mudanças um pouco mais subtis, que têm a ver com a vulgarização do uso dos computadores muito para além dos ambientes especializados em que se implantaram inicialmente. Do uso dos computadores em grandes organizações para tarefas de processamento de grandes volumes de dados e em departamentos especializados de universidades para investigação passou-se, a partir dos anos 70 com os micro-computadores, à sua introdução em tarefas mais diversificadas e em organizações de menores dimensões e rapidamente à sua presença na vida de todos os dias com a generalização da introdução dos computadores tanto no ambiente das organizações como no doméstico [CBI 2003].

Uma nova dimensão no uso dos computadores surgiu com o aparecimento das

ligações em rede e com a criação da Web. Para além das ferramentas disponíveis em cada computador, os utilizadores passaram a ter acesso a informação das mais variadas fontes, através de uma interface acessível sem conhecimentos técnicos. Tornou-se fácil trocar informação em formato máquina entre utilizadores não especializados e portanto produzir novos documentos com base na informação existente.

Do ponto de vista das Ciências Documentais, esta mudança traz novas perspectivas para as tarefas tradicionais mas requer também a consideração de uma nova realidade no tratamento e acesso aos documentos. Por um lado, criou-se nos utilizadores dos recursos em linha uma grande expectativa quanto à acessibilidade dos recursos e a disponibilidade para a sua consulta sem intermediários. Por outro lado, a criação e a disponibilização de enormes volumes de informação na Web coloca questões complexas na sua validação, evolução e preservação.

Uma nova vaga de mudanças tecnológicas está em vista com a actividade que se desenvolve actualmente no âmbito do consórcio W3C sob a designação de Web semântica. O objectivo da Web semântica é tornar a informação na Web inteligível não apenas para as pessoas, seus utilizadores finais, mas também para as máquinas, que cada vez mais são intermediários para as tarefas complexas que os utilizadores pretendem realizar sobre a informação disponível. A Web semântica pretende fazer evoluir a Web de uma rede em que a informação é facilmente trocada entre máquinas mas o seu significado apenas acessível às pessoas para aquela em que o significado dos dados possa também ser manipulado pelas máquinas.

A tarefa de construir uma Web dos significados, embora requeira o desenvolvimento de novas soluções tecnológicas para normalizar as representações dos significados, depende crucialmente da capacidade de elaborar modelos adequados para a representação dos documentos, tarefa central para as Ciências Documentais e da Informação.

A Pesquisa de Informação

A área da pesquisa de informação, condicionada tradicionalmente pela capacidade de tratamento manual da informação, desenvolveu-se a partir dos anos 50 no sentido da automatização. Os progressos nesta área [Sparck-Jones 2003] mostraram que a indexação e pesquisa automáticas, mais do que mimar as técnicas

usadas manualmente, abriram caminho a novos modelos e técnicas com resultados que excedem os obtidos pelos métodos tradicionais, para além de permitirem a experimentação numa escala impraticável para estes.

Dá-se por isso a coincidência, nos anos 90, de a criação da Web surgir num tempo em que os métodos de indexação e pesquisa automática tinham já provado o seu valor. A pesquisa na Web forneceu-lhes um campo de experimentação de outra ordem de grandeza, ao mesmo tempo que permitiu explorar aspectos da informação inexistentes nos meios tradicionais, como é o exemplo das ligações no hipertexto, correntes nos documentos na Web.

As limitações presentes na pesquisa de informação situam-se nas áreas onde as barreiras no acesso ao significado dos documentos são maiores. Exemplos disto são a informação visual - imagens e vídeo - e a informação multi-língua. Nestes domínios a investigação em métodos automáticos tem sido intensa mas os resultados estão ainda longe dos obtidos na pesquisa de informação textual no que diz respeito à satisfação das necessidades dos utilizadores.

A área da pesquisa de informação é assim uma das que pode beneficiar com a evolução da descrição da informação, podendo tirar partido do enriquecimento dos documentos disponíveis e das possibilidades abertas para a interpretação automática dos seus conteúdos.

A produção de documentos electrónicos

Os contextos de produção de documentos tradicionais beneficiavam do facto de os documentos surgirem enquadrados em processos bem definidos dentro das organizações e de ser possível a sua gestão nesse contexto.

A mudança para a produção de documentos com suporte em aplicações informáticas introduziu modificações em muitos aspectos, de que se vão destacar dois.

Usando como exemplo a informação de natureza contabilística de uma organização, pode-se dizer que a sua produção em papel fornecia uma visão fácil de interpretar para um utilizador humano, corporizando-se em documentos produzidos a diversos níveis da organização, relacionados com os seus processos de funcionamento e dando origem em pontos especificados aos documentos requeridos oficialmente. Quando a organização passa a gerir a sua contabilidade usando uma aplicação

informática suportada numa base de dados, ainda que os processos desenvolvidos sejam os mesmos, a informação contabilística passa a estar estruturada segundo um modelo interno à base de dados. Alguns dos modos de uso e pesquisa nessa informação deixam de ser baseados em documentos. O arquivo e preservação desta informação coloca problemas que estão ainda em aberto, desde saber o que constitui a informação a preservar até ao estabelecimento das normas e dos formatos necessários à sua manipulação [DPC 2002].

Um outro aspecto na gestão da informação a nível das organizações é o facto de, para além da informação que reside nos sistemas de informação e que é objecto de particular cuidado por dizer respeito ao núcleo das suas actividades, todo o pessoal da organização produzir os seus próprios documentos - para comunicação interna, para contacto com parceiros, via correio electrónico, em relatórios de divulgação mais ou menos generalizada. Esta informação tem correntemente uma natureza de organização “ao documento” mas, sendo produzida por ferramentas muito diversas, sofre de problemas de gestão óbvios: é difícil estruturá-la a nível da organização e está frequentemente representada em formatos privativos de aplicações que são descontinuados e geram a perda efectiva dos conteúdos.

Em ambos os casos ilustrados a introdução de uma nova sofisticação no armazenamento e tratamento da informação não foi acompanhada do desenvolvimento de modelos que permitam interpretá-la fora do contexto estrito das operações que suporta, e portanto fomentar o seu uso mais flexível e a sua preservação.

A informação na Web

A vulgarização do uso dos computadores e da acessibilidade à Web produziu uma mudança nos modos de uso da informação cuja extensão ainda não está completamente à vista. Para o utilizador leigo, a navegação na Web põe à vista uma quantidade de informação avassaladora, criando por um lado a impressão de que tudo o que se pode querer consultar está lá, e por outro alguma sensação de impotência pelo facto de, como resultado de uma consulta realizada usando um motor de busca, haver aparentemente uma quantidade não manejável de documentos na resposta obtida.

Do ponto de vista do fornecimento de informação, a Web criou nos utilizadores

uma expectativa muito grande da disponibilidade da informação que tradicionalmente requeria um contacto mais personalizado ou a intervenção de um técnico. Passou a esperar-se das empresas, das instituições públicas e dos mais diversos serviços que tivessem a sua informação consultável na Web. Do lado das organizações a Web é vista como um meio poderoso de acesso às pessoas e muitos dos processos de interacção com os utilizadores estão a ser substituídos, com vantagens para ambos os intervenientes, por serviços prestados através da Web.

Do ponto de vista das organizações cuja missão já era o fornecimento de informação- bibliotecas, serviços de documentação, arquivos, a era da Web veio criar a necessidade de fornecer na Web e directamente aos utilizadores finais serviços menos especializados do que os que eram prestados por técnicos conhecedores do domínio. Por outro lado as facilidades de comunicação permitem alargar o âmbito destes serviços, oferecendo uma vista integrada da informação de diversas fontes e uma efectiva separação entre a localização física dos recursos e a sua integração lógica.

Internamente às organizações, os conceitos introduzidos pela Web também produziram mudanças na estruturação das aplicações de suporte ao funcionamento. Onde antes eram comuns aplicações dedicadas a tarefas específicas de um departamento passa agora a haver diferentes interfaces para um repositório de informação comum que, conforme o departamento e o utilizador, dão acesso às funções relevantes ou autorizadas. O facto de ser adoptado o padrão de interacção na Web separa os requisitos de acesso dos utilizadores das características do posto de trabalho: qualquer computador da organização pode oferecer a qualquer utilizador acesso às funcionalidades requeridas pelas suas funções.

Genericamente e para aplicações que são projectadas para o uso em diversos ambientes (utilitários, acesso a recursos) está a tornar-se norma a interface da aplicação ser projectada para uso através do “browser” Web. Para aplicações que requerem a manipulação de documentos, as interfaces tiram em geral partido das facilidades do hipertexto, sendo portanto cada vez mais comum a reconstrução de documentos explicitando a sua estrutura e referências.

A massificação da disponibilização e do uso da informação na Web tem feito reflectir sobre a qualidade da informação disponível. Nos modos de edição tradicionais a informação passava por um processo que garantia ao utilizador um grau de confiança associado à credibilidade das instituições de edição. No ambiente da

Web a facilidade de publicação leva a que, mesmo da parte de instituições com grande credibilidade, sejam disponibilizados conteúdos de natureza provisória a par com os que correspondem aos parâmetros de qualidade da edição tradicional. Tem havido mesmo, por parte de organizações científicas, a prática de tornar os artigos submetidos para publicação visíveis na Web, estando abertos a comentários antes da sua publicação definitiva. Esta é uma forma de tirar partido das facilidades de comunicação para tornar mais vivo e aberto o processo de edição. Este modo de funcionamento tem no entanto como reverso a possibilidade de se disseminarem documentos cujo estado é difícil de atestar. Mais genericamente, qualquer pessoa ou organização pode publicar informação cujo conteúdo é de alguma forma objectável; para os mecanismos de pesquisa actualmente disponíveis esta pode não ser distinguida da que provém de uma fonte conceituada, e é o utilizador que tem de ter os conhecimentos necessários para fazer a distinção.

Comunicação e Significado da Informação

A troca de informação entre pessoas e o seu uso nas mais diversas actividades supõe um entendimento quanto ao significado da informação manipulada. Entre pessoas a comunicação efectiva depende da partilha de um contexto que forneça significado à informação. É muitas vezes necessário explicitar o significado de conceitos que são usados implicitamente na comunicação humana, seja porque é necessário restringir num determinado domínio o significado de um termo relativamente ao seu uso na linguagem comum, seja porque se pretende evitar ambiguidades, seja ainda porque se pretende fornecer ao interlocutor (ou mais genericamente a alguém que vai usar a informação) o contexto exacto em que a deve interpretar. Esta explicitação é tipicamente feita em linguagem natural, de uma maneira rigorosa mas informal, sendo fácil de utilizar por pessoas.

Quando se passa da comunicação entre pessoas para a comunicação entre aplicações, a linguagem deixa de poder ser informal. Uma aplicação tem habitualmente um domínio cujos conceitos são captados tanto na informação que é manipulada como nas funcionalidades oferecidas ao utilizador da aplicação. Pode-se dizer que a semântica do domínio está implicitamente captada na aplicação. Quando há necessidade de comunicação entre aplicações, é habitual estabelecer uma linguagem própria, que mais uma vez tem implícitos os conceitos do domínio. Estas

linguagens, habitualmente designadas protocolos, são faladas pelas aplicações quando precisam de trocar informação entre si. A informação trocada tem significado para as aplicações, e pode ser transformada por estas em informação inteligível para o utilizador humano.

No desenvolvimento de aplicações informáticas de alguma dimensão é habitual fazer a distinção entre várias “camadas”. A informação manipulada pela aplicação é gerida por um servidor de dados, que trata do armazenamento e da resposta aos pedidos da aplicação, ao passo que o servidor da aplicação realiza sobre os dados os procedimentos necessários para cumprir as funcionalidades da aplicação. Sobre estas camadas situa-se a interface do utilizador, que dá acesso às funções da aplicação adaptadas a um tipo de utilizador e a um ambiente. Este modelo em camadas favorece uma visão em que os dados estão separados das operações que sobre estes se realizam e da forma de as apresentar ao utilizador. Um mesmo servidor de dados pode suportar aplicações diferentes sobre um mesmo domínio, e uma mesma aplicação pode apresentar-se de forma diversa a utilizadores que lhe acedam de diferentes plataformas ou que disponham de privilégios diferentes.

Neste modelo é claro que o significado da informação está implícito no modelo de dados usado pelo servidor de dados, e que este tem de ser partilhado pelo servidor da aplicação que a lê, escreve e interroga.

A Web Semântica: Explicitar o Significado da Informação

As linguagens usadas na Web foram desenhadas com o objectivo de serem de acesso fácil para pessoas. A maior parte da informação disponível na Web encontra-se em formato de texto com hiperligações, tornando fácil a navegação entre documentos. A exploração intensa da Web fez aumentar o seu uso por pessoas mas fez também surgir o conceito de “serviços”, ou seja de aplicações que se destinam a realizar tarefas em nome do utilizador, explorando por ele a informação disponível e fornecendo um resultado que ele não poderia obter ou que lhe consumiria demasiado tempo.

O que distingue uma aplicação que opera sobre informação disponível na Web para realizar uma determinada tarefa de uma aplicação que usa uma base de dados de grande dimensão para uma tarefa semelhante? A diferença essencial está na falta de conhecimento do modelo dos dados no caso da Web. Se queremos uma aplicação que

opera sobre informação proveniente de fontes muito heterogêneas, temos de deixar de a basear num modelo de dados implícito.

Se analisarmos a manipulação da informação desde a comunicação entre as pessoas em linguagem natural, em que a semântica é informal e implícita, passando pela sua explicitação na descrição informal de conceitos, e evoluindo para uma captação mais estrita e menos ambígua nos modelos de dados partilhados entre uma aplicação e o seu servidor, observamos parte de um “contínuo semântico” [Ushold 2003] que até este ponto pode já ter semântica expressa formalmente mas a que falta ainda a possibilidade de a processar automaticamente nas aplicações. Este último nível é o que deve ser oferecido no ambiente da “Web Semântica”.

Segundo Tim Berners-Lee, criador da Web e impulsionador da iniciativa da Web Semântica no consórcio W3C, “A Web Semântica é uma extensão da Web actual na qual a informação recebe um significado bem definido, permitindo que os computadores e as pessoas cooperem melhor”[Berners-Lee et al. 2001].

No estado actual do desenvolvimento, a Web Semântica tem, para além deste propósito genérico, diversas tecnologias desenvolvidas e em desenvolvimento para o concretizar. As mais básicas estão ao nível da uniformização da sintaxe usada para representar dados elementares, em que as linguagens de anotação da família do XML [W3C] se estão já a tornar a norma de facto. Este primeiro nível garante um formato aberto e a possibilidade de criar anotações arbitrárias, ou seja completa liberdade em termos de criação de conceitos.

A Web Semântica pode ser vista como o alargamento para a Web dos esforços desenvolvidos na Ciência da Computação, desde os anos 50, no sentido de representar conhecimento de uma forma adequada à automatização do raciocínio [Brachman et al. 1985, Sowa 2000]. O primeiro aspecto que teve de ser visto a uma nova luz no âmbito da Web foi o da abertura em termos de criação de conceitos: ao contrário do que poderia ser feito na modelização de um domínio particular (embora mesmo neste caso a tarefa possa não ser fácil) é impensável na Web centralizar modelos. É preciso permitir que cada comunidade crie os seus próprios conceitos e os possa exprimir sem limitações. Esta facilidade é proporcionada pela uniformidade sintáctica das linguagens de base.

As aplicações actuais já fazem uso das anotações presentes nos documentos e que pretendem explicitar a sua estrutura e o significado das suas componentes. Este uso assume, no entanto, o conhecimento da semântica dessas anotações, que pode ser

inconsistente entre o criador dos dados e a aplicação que os usa. A anotação em XML permite assim que cada utilizador estruture livremente os seus documentos, mas não estabelece nada quanto ao significado dessa estrutura.

O nível seguinte nas tecnologias para a Web Semântica é o RDF ou “Resource Description Framework” [W3C 2003], que é uma linguagem para fazer em XML afirmações acerca de recursos. As “frases” RDF são termos com um sujeito, um verbo e um objecto e permitem formular afirmações como “X é o autor de Y” em que X e Y podem corresponder a recursos na Web (identificados por URI’s, “Universal Resource Identifiers” [W3C 2003]), pessoas ou o que se quiser, e a propriedade “é o autor de” pode ela própria estar definida num URI. A ligação de recursos e conceitos a URI’s é fundamental aqui, uma vez que permite associar os conceitos usados nas afirmações RDF às suas definições situadas num local único na Web.

As Ontologias

O uso do RDF permite escrever afirmações que captam o significado dos dados numa aplicação. Se cada aplicação pode criar os seus próprios conceitos, e pretendemos que outras possam manipulá-los de uma forma sensível ao significado, então é preciso que o significado desses conceitos seja exposto de uma forma que se possa manipular automaticamente. Na Web Semântica esse é o papel das ontologias [W3C 2003, DAML 2003, Hendler et al. 2002].

O termo “ontologia”, de uma área da Filosofia, foi recentemente adoptado na Ciência da Computação com um significado mais estrito que tem servido para identificar no âmbito desta comunidade os conceitos formalizáveis desta área, na perspectiva da computação. Na Filosofia, a ontologia questiona o ser e a existência, investigando a natureza, propriedades essenciais e relações de todos os seres, e os próprios princípios e causas da existência [Hyperdictionary 2003]. Na Ciência da Computação, e mais especificamente nos domínios da Inteligência Artificial onde começou por ser adoptado, o termo tem servido para caracterizar uma especificação formal e explícita de como representar objectos, conceitos e outras entidades cuja existência é assumida numa determinada área, bem como as relações que se verificam entre estes.

Este uso estrito do termo, embora conforme ao seu espírito mais abrangente, limita-se aos seus aspectos “automatizáveis”, e nesse sentido uma definição corrente

para ontologia é “a especificação de uma conceptualização”[Gruber 2003]. Na sua versão mais elementar, uma ontologia pode reduzir-se a uma hierarquia taxonómica, com a relação de subclasse associada. Genericamente uma ontologia pode incluir as definições que associam nomes de entidades num domínio de discurso (classes, relações, funções, ou outros) com descrições informais do seu significado e os axiomas que restringem as interpretações e o uso de tais nomes. Neste sentido uma ontologia pode ser usada como uma teoria lógica.

Em termos de uso, a ontologia serve de vocabulário para a troca de informação entre aplicações. A adopção de uma ontologia é um compromisso de consistência com os seus conceitos, e a garantia de que uma aplicação, dentro dos limites das suas funcionalidades, irá interrogar outras aplicações e responder aos pedidos destas de acordo com o vocabulário comum. Dentro do modelo aberto da Web estas aplicações podem situar-se a níveis diferentes no uso da informação, terem objectivos diversos e lidarem com informação mais ou menos complexa. A partilha da ontologia pretende garantir consistência na informação trocada, mas não completude: um serviço pode ser incapaz de responder a um pedido que lhe é enviado, apesar de este estar formulado segundo conceitos da ontologia a que conforma [Gruber 2003].

Papel da Ciência da Informação

Observando o panorama das tecnologias relacionadas com a manipulação da informação verificamos que estas se estendem por áreas que vão dos sistemas de informação e bases de dados às redes, das tecnologias Web à representação de conhecimento e à inferência automática. Este universo está em evolução tão rápida que é impossível, mesmo para quem trabalha dentro das tecnologias, manter-se a par delas em todas as suas dimensões. Torna-se por isso difícil ver quais os aspectos que os especialistas das Ciências Documentais e da Ciência da Informação podem influenciar com os seus conhecimentos e visão dos problemas.

O primeiro cenário que se pode imaginar é bastante conservador, e reflecte de certo modo o que se tem visto no passado recente com a introdução das chamadas tecnologias da informação nos domínios tradicionais das bibliotecas e da documentação. Do lado dos especialistas nestas áreas nota-se um grande desajuste relativamente à aplicação das tecnologias. Acontece frequentemente que o especialista da informação não se apercebe das possibilidades oferecidas pela tecnologia e é pouco

ambicioso nos seus projectos. Possivelmente ocorrerá com a mesma facilidade a situação em que da tecnologia se esperam soluções para problemas que requerem mais do que tecnologia. Em qualquer dos casos não é fácil, nem para o especialista da informação nem para o das tecnologias, abarcar toda a complexidade do domínio complementar. Se se continuar nesta linha será de esperar maiores avanços e resultados mais visíveis nas áreas onde as tecnologias estão bem implantadas, como as bibliotecas, e a continuação de um certo afastamento naquelas onde os investigadores da informação estão menos inseridos na cadeia de fornecimento da informação, como é o caso dos arquivos. Neste cenário, à medida que as ontologias surgem e são refinadas será necessário incorporá-las nos novos mecanismos de estruturação e arquivo da informação.

Num cenário mais optimista podemos ver a contribuição dos especialistas da informação a surgir inserida nos próprios processos de definição da estrutura da informação, do seu significado e dos seus modos de uso. São claras as áreas onde esta contribuição pode ser significativa. Enumeram-se de seguida algumas.

- Modelização de sistemas de informação - nas fases de definição dos conceitos do domínio, da natureza dos dados e das funcionalidades das aplicações suportadas num sistema de informação a colaboração entre os especialistas da informação, os das tecnologias e os das aplicações permite embutir nos modelos desenvolvidos as facetas relevantes tanto para as suas funções de suporte às aplicações como para as de repositório multifacetado.
- Desenvolvimento de modelos para pesquisa de informação - trata-se aqui de contribuir para construir modelos que usem a informação nas suas diversas formas e que sejam adequados a cumprir a função de recuperação de informação usando toda a flexibilidade oferecida pela exploração da semântica dos dados.
- Desenvolvimento de interfaces para aplicações e serviços de troca de informação - pretende-se aqui fazer uso da sensibilidade dos profissionais da informação aos modos de interacção e de consulta de informação e à sua evolução.
- Definição, implementação e acompanhamento dos processos de preservação de informação em qualquer formato e meio - esta, embora seja uma tarefa pouco visível, é das que mais se modificou com os novos meios de produção de informação. Está-se neste momento numa situação em que não é praticável o

arquivo *a posteriori* de uma parte apreciável dos documentos que são produzidos, sendo condicionante da preservação a intervenção dentro do processo de criação. A contribuição dos especialistas da informação é aqui essencial, uma vez que, sendo a preservação e arquivo funções vistas como menos prioritárias dentro das aplicações, facilmente são deixadas fora do desenho inicial, e dificilmente surgem de forma limpa na continuação do desenvolvimento.

Em suma, o conhecimento sobre a descrição de documentos é valioso tanto no desenvolvimento de ontologias, como na especificação dos padrões de inferência válidos para os documentos. Pode ainda contribuir para a especificação das componentes de arquivo das aplicações para gestão de informação, bem como para as regras de autenticação dos documentos.

As ciências ligadas aos domínios das humanidades têm uma tradição de evolução muito ponderada, preferindo discutir em profundidade os conceitos antes de avançar para explorar as suas ramificações na prática profissional. O acompanhamento das tecnologias vai requerer um tipo de atitude mais “experimental”, avançando com demonstrações de conceitos ainda antes de estes serem aceites e validados pela comunidade. Sem isto, há um risco efectivo de que os avanços se façam apesar da imaturidade dos conceitos, e que depois não reste senão aceitar normas para as quais os profissionais da área não puderam contribuir.

Conclusões

Nas reflexões recentes, da parte de investigadores oriundos das Ciências Documentais, sobre a construção científica da área da Ciência da Informação [Silva e Ribeiro 2002], são visíveis o aprofundamento do seu objecto de estudo e o estabelecimento dos métodos de investigação apropriados. O lançamento das bases para a Ciência da Informação trará decerto uma maior sensibilidade a problemas que vão para além dos tradicionais nas Ciências Documentais e uma ligação mais forte aos temas da Informática.

Pretendeu-se aqui fazer uma visita aos domínios da “Web Semântica”, fazendo sobressair tanto a sua abrangência em termos tecnológicos como as oportunidades que existem, nas áreas que estão a ser exploradas, para os investigadores da Ciência da Informação.

Vê-se frequentemente a sofisticação tecnológica ser imposta às pessoas sem que isso seja acompanhado de uma efectiva apropriação das novas ferramentas por parte dessas pessoas. A tecnologia surge assim mais como uma barreira ao aprofundamento das capacidades humanas do que como uma chave de acesso a novas vistas sobre o mundo e novas formas de interacção.

Nesta área iremos observar decerto experiências mais e menos bem sucedidas na aplicação da tecnologia aos problemas do dia-a-dia. É bem certo que para alguns dos problemas centrais da nossa relação com a informação ainda não existem soluções definitivas. Alguns destes são no entanto bem familiares aos investigadores das Ciências Documentais. Ao adoptarem uma estratégia de abertura para as possibilidades oferecidas pelas tecnologias, estes irão ter acesso a ferramentas que lhes podem oferecer meios de criar soluções, testá-las e partilhá-las. A experiência acumulada tanto em termos de problemas genéricos como de casos concretos de aplicação será valiosa tanto para sugerir novas soluções para problemas recorrentes como para criar soluções para os novos problemas que a evolução tecnológica cria.

Referências

- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora - Semantic Web. *Scientific American*. 284:5 (2001).
- BRACHMAN, Ronald J.; LEVESQUE, Hector, J. - *Readings in knowledge representation*. Los Altos, Cal: Morgan Kaufmann, 1985.
- CBI - Charles Babbage Institute- Center for the History of Information Technology [Em linha]. Actualizado 1 Ago. 2003. [Consultado 17 Out. 2003]. Disponível em <http://www.cbi.umn.edu/shp/bibliography.html>.
- DAML - The DARPA Agent Markup Language Homepage [Em linha]. Actualizado 10 Jun. 2003. [Consultado 17 Out. 2003]. Disponível em <http://www.daml.org/>.
- DPC - The Digital preservation coalition [Em linha]. Actualizado 2002. [Consultado 17 Out. 2003]. Disponível em <http://www.dpconline.org/graphics/index.html>.
- GRUBER, Tom - What is an ontology? [Em linha]. [Consultado 17 Out. 2003]. Disponível em <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- HENDLER, James; BERNERS-LEE, Tim; MILLER, Eric - Integrating

applications on the semantic Web. *Journal of the Institute of Electrical Engineers of Japan*. 122:10 (2002) 676-680. Disponível em <http://www.w3.org/2002/07/swint>.

- Hyperdictionary [Em linha]. [Consultado 17 Out. 2003]. Disponível em <http://www.hyperdictionary.com/dictionary/ontology>.
- SILVA, Armando Malheiro da; RIBEIRO, Fernanda - *Das “Ciências” Documentais à Ciência da Informação*. Porto: Edições Afrontamento, 2002.
- SOWA, John F. - *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove, CA: Brooks Cole Publishing, 2000.
- SOWA, John F. - Ontology. [Em linha]. Atualizado 16 Jun. 2000. [Consultado 17 Out. 2003]. Disponível em <http://users.bestweb.net/~sowa/ontology/index.htm>.
- SPARCK-JONES, Karen - Document retrieval : shallow data, deep theories, historical reflections, potential directions. In ECIR 2003 - EUROPEAN CONFERENCE ON INFORMATION RETRIEVAL RESEARCH, 25 - *Lecture Notes in Computer Science 2633*. p. 1-11.
- USCHOLD, Michael - Where are the semantics in the semantic web?. *AI Magazine*. 24:3 (2003) 25-36.
- W3C - Extensible Markup Language (XML) [Em linha]. Atualizado 20 Ago. 2003. [Consultado 17 Out. 2003]. Disponível em <http://www.w3.org/XML/>.
- W3C - Naming and addressing: URIs [Em linha]. Atualizado 9 Jul. 2002. [Consultado 17 Out. 2003]. Disponível em <http://www.w3.org/Addressing/>.
- W3C - RDF Primer [Em linha]. Atualizado 10 Out. 2003. [Consultado 17 Out. 2003]. Disponível em <http://www.w3.org/TR/rdf-primer/>.
- W3C - Semantic Web [Em linha]. Atualizado 11 Out. 2003. [Consultado 17 Out. 2003]. Disponível em <http://www.w3.org/2001/sw/>.
- W3C - Web-Ontology (WebOnt) Working Group [Em linha]. Atualizado 15 Out. 2003. [Consultado 17 Out. 2003]. Disponível em <http://www.w3.org/2001/sw/WebOnt/>.