

Organização da informação em Subject Gateways

Paulo Jorge Leitão

Orientador: Dr^a Fernanda Maria Campos
Co-Orientador: Prof. Dr. Pedro Faria Lopes

1

Organização da informação em Subject Gateways

I – Tema. Objectivos. Metodologia

1.1. Tema

Subject Gateways sob o ponto de vista dos processos e formas de organização da informação

1.1. *Subject Gateways*: definição

Serviços de acesso à Internet (Web) para a comunidade universitária, caracterizando-se por:

- a) Definirem e implementarem uma rigorosa política de selecção de recursos Internet, maioritariamente de acesso livre, adequados à comunidade que pretendem servir, pautada por critérios de qualidade, autoridade, actualidade e relevância do assunto. Os recursos não são apropriados por este tipo de serviços;

2

Organização da informação em Subject Gateways

- b) Desenvolverem metodologias de gestão da colecção que podem passar pela remoção de recursos que se tornaram desapropriados, verificação da validade das ligações, reavaliação periódica da adequação do recurso aos objectivos do serviço.
- c) Descreverem os recursos seleccionados através de linguagens de metadados estruturadas e de conteúdo informativo rico, em formatos com campos individuais
- d) Utilizarem linguagens documentais, nomeadamente classificações, para a representação do conteúdo, o que permite criar estruturas que facilitam o *browsing* dos recursos.

3

Organização da informação em Subject Gateways

1.2. Objectivos:

- a) Analisar as formas de organização da informação desenvolvidas por estes serviços a partir do estudo de um caso paradigmático: os primeiros *Gateways* criados no âmbito da RDN (Resource Discovery Network)
- b) Avaliar a eficácia das formas de organização de informação implementadas na prossecução dos objectivos a atingir por qualquer sistema de informação de tipo bibliográfico.

4

Organização da informação em Subject Gateways

1.3. Metodologia

1.3.1. Opção metodológica principal: estudo de caso.

Caso: 5 *Gateways* iniciais da RDN:

- SOSIG: social science information gateway
- HUMBUL: humanities hub
- PSIGATE: the physical sciences information gateway
- EEVL: the Internet guide to engineering, mathematics and computing
- BIOME: your guide to Internet resources in the health and life sciences.

5

Organização da informação em Subject Gateways

1.3.2. Percorso metodológico

- a) Análise da literatura relevante sobre a Web em ordem a caracterizar: tipos de informação existente e tipos de serviços de recuperação
- b) Análise da literatura relevante sobre organização da informação, quer no que diz respeito ao modelo tradicional quer quanto a metadados
- c) Revisão da literatura sobre *Subject Gateways*
- d) Análise das construções teórico-práticas específicas: Desire Gateways Handbook, ROADS Cataloguing Guidelines (*Resource Organization and Discovery in Subject-Based Services*)

6

Organização da informação em Subject Gateways

e) Definição de critérios de análise e avaliação das formas de organização da informação implementadas pelos serviços, definidos a partir do modelo de organização da informação de tipo bibliográfico. A avaliação foi realizada a dois níveis:

- Ao nível das concepções plasmadas nos normativos da descrição criados para cada serviço
- Ao nível da prática, através da análise de uma amostra de 50 descrições de cada serviço e das estruturas de organização do *browsing* dos recursos

7

Organização da informação em Subject Gateways

II – Estudo do caso

2.1. Linguagens da descrição

- Definição de um normativo para a descrição em cada serviço: Regras de Catalogação
- Linguagens privilegiam a descrição da Obra enquanto entidade intelectual
- Descrição do documento, enquanto invólucro físico, preocupa-se sobretudo com a acessibilidade
- Definição clara das entidades a descrever: obra, documento, autores, assuntos
- Vocabulário claramente definido, mas sobretudo derivado da fonte prescrita de informação. Reduzido controlo vocabular, exceptuando o caso da representação do conteúdo

8

Organização da informação em Subject Gateways

II – Estudo do caso

2.1. Linguagens da descrição

- Atributos das entidades a descrever inequivocamente definidos
- Reduzido tratamento do atributo Relação
- Semântica e sintaxe pouco desenvolvidas, exceptuando-se o caso dos assuntos

2.2. Prática da descrição

- Na globalidade, consistência e uniformidade na aplicação dos normativos
- Reduzida taxa de erros
- Tratamento menos consistente do atributo Autor.

9

Organização da informação em Subject Gateways

II – Estudo do caso

2.3. Classificação e Indexação

- Utilização de linguagens normalizadas, com origem no paradigma biblioteconómico tradicional
- Classificações, Tesouros, Listas de Descritores
- Linguagens diferenciadas para desempenhar funções diferentes:
 - Classificações: representação do conteúdo a um nível de generalidade, tendência contrariada apenas por EEVL e SOSIG
 - Tesouros: representação do conteúdo com grande especificidade
- Ausência de qualquer forma de relação entre as linguagens
- Aplicação consistente das linguagens de acordo com as funções para que são utilizadas

10

Organização da informação em Subject Gateways

II – Estudo do caso

- Atributo "Categoria / Tipo de recurso": desempenha funções de qualificador de forma

2.4. *Browsing*

- Todos os serviços implementam estruturas de organização de conteúdos que permitem o *browsing* dos recursos, através da utilização das classificações e/ou tesouros
- Classificações: representação generalista, adaptando-se a um comportamento de pesquisa semi-estruturado
- Tesouros: representação específica, adaptando-se a um comportamento de pesquisa estruturado

11

Organização da informação em Subject Gateways

II – Estudo do caso

2.4.1. Classificações

- Implementação: modelo de organização hierárquico ou alfabético
- Flexibilidade na utilização das classificações: posição superior de classes representadas por notações mais específicas
- Relação entre nível de especificidade da representação e o número de recursos classificados.
- Alta taxa de atribuição em quase todos os serviços

2.4.2. Tesouros

- Implementação: modelo alfabético

12

Organização da informação em Subject Gateways

II – Estudo do caso

2.5. Descrições em texto livre

- Tratamento desenvolvido deste atributo
- Aspectos mais desenvolvidos: cobertura temática do recurso, responsabilidade e tipologia
- Ausência de definição do estilo do resumo: informativo, indicativo ou crítico
- Ausência de aspectos essenciais de caracterização dos documentos neste ambiente: estrutura dos sites web, formas de navegação, entre outras.
- Perfil dos resumos produzidos: textos não estruturados, extensão média, frases de alguma complexidade mas utilizando palavras simples, grande utilização da voz passiva, reduzido nível de legibilidade

13

Organização da informação em Subject Gateways

III – Conclusões

- Modelo de organização da informação baseado em linguagens de metadados estruturadas e com conteúdo relativamente rico, privilegiando a descrição da Obra.
- Clara predominância dos atributos que representam o conteúdo: Assuntos, Palavras-Chave, Descrições em texto livre. Predomínio do controlo vocabular.
- Descrições em texto livre: necessidade de repensar uma semântica descritiva mais adaptada ao tipo de recurso e ambiente em que são disponibilizados

14

Organização da informação em Subject Gateways

III – Conclusões

- Privilégio do ponto de acesso Assunto, particularmente através das estruturas de *browsing*. Implementação flexível das linguagens documentais. Modelo hierárquico ou alfabético apresenta dificuldades no crescimento.
- Atributos maioritariamente coincidentes entre as várias linguagens: Título próprio, Complemento de título, Título paralelo, Autor pessoa física, Descrição em texto livre, Assunto (classificação), Palavras-chave, Língua, Tipo de recurso, URL.
- Raros dispositivos de controlo semântico e sintáctico, exceptuando o caso da representação do conteúdo
- Deficiente tratamento dos atributos Autor e Relação. Necessidade de pensar uma taxonomia das relações entre recursos neste ambiente

15

Organização da informação em Subject Gateways

III – Conclusões

- Nível de consecução dos objectivos de um sistema de informação bibliográfico:
 - Alta probabilidade de concretização dos seguintes objectivos: encontrar uma entidade singular, localizar todos os documentos com o mesmo assunto, identificar e seleccionar uma entidade apropriada às necessidades do utilizador, obter acesso à entidade descrita
 - Objectivos de difícil concretização: localizar todos os documentos da mesma Obra, localizar todos os documentos da mesma versão, localizar todos os documentos de um mesmo autor, navegabilidade no universo bibliográfico

16

Organização da informação em Subject Gateways

IV - Questões estratégicas para reflexão

1. Diminuição dos custos de produção e manutenção do serviço

- 1.1. Hipótese 1: redução da complexidade da descrição. Desvantagem: perda significativa de qualidade para o utilizador final
- 1.2. Hipótese 2: Geração assistida de metadados
 - 1.2.1. Geração automática de metainformação descritiva. Exemplo DC-dot (www.ukoln.ac.uk/metadata/dcdot).
 - 1.2.2. Classificação / indexação automática
- 1.3. Hipótese 3: "Harvesting" metadata
 - 1.3.1. Protocolo OAI (Open Archives Initiative), por exemplo (www.openarchives.org)

17

Organização da informação em Subject Gateways

2. Alargamento do espectro de possibilidades de pesquisa no mesmo contexto

- 2.1. Hipótese 1: Alargamento da abrangência temática do serviço
 - 2.1.1. Solução HUB, seguida pelos *Gateways* da RDN
- 2.2. Hipótese 2: Assegurar a pesquisa cruzada de múltiplos *Gateways*. Problema: interoperabilidade entre linguagens de metadados
 - 2.2.1. Normalizar uma linguagem mínima comum, a ser tendencialmente utilizada por todos. Exemplo: RDN Cataloguing Guidelines
 - 2.2.2. Mapeamento de linguagens documentais diversas para uma linguagem comum. Perspectiva do serviço RENARDUS (www.renardus.org): mapeamento de classificações dos serviços participantes para CDD (classificação decimal de Dewey)

18