

Mapeamento de Domínios do Conhecimento

Técnicas e Resultados

Rosane Minghim

Workshop de Mapeamento Visual de Coleções de Documentos – Set 2005
ICMC – Universidade de São Paulo



ORGANIZAÇÃO

- MOTIVAÇÃO
- INTRODUÇÃO
 - Visualização e Mineração Visual
 - Visualização de Domínios do Conhecimento
 - *Visual Analytics*
- TÉCNICAS
 - Trabalhos relacionados
 - Nossa abordagem
- RECURSOS
- PANORAMA

Workshop de Mapeamento Visual de Coleções de Documentos

ORGANIZAÇÃO

- **MOTIVAÇÃO**
- INTRODUÇÃO
 - Visualização e Mineração Visual
 - Visualização de Domínios do Conhecimento
 - *Visual Analytics*
- TÉCNICAS
 - Trabalhos relacionados
 - Nossa abordagem
- RECURSOS
- PANORAMA

Workshop de Mapeamento Visual de Coleções de Documentos

Motivação

- DISPONIBILIDADE DE INFORMAÇÃO NA FORMA TEXTUAL
 - Resultados de busca
 - Sítios de informação bibliográfica
 - Bancos de Dados de Documentos
 - Comunicação Humana
 - ...
- DIFICULDADES DE INTERPRETAÇÃO
- APLICAÇÕES (demanda)
 - Pesquisa Científica
 - Pesquisa Jornalística
 - Investigação
 - Atendimento a Clientes/Usuários
 - **Organização e Localização de Material**
 - Acompanhamento de Evolução de Assunto

Workshop de Mapeamento Visual de Coleções de Documentos

ORGANIZAÇÃO

- MOTIVAÇÃO
- **INTRODUÇÃO**
 - **Visualização e Mineração Visual**
 - **Visualização de Domínios do Conhecimento**
 - *Visual Analytics*
- TÉCNICAS
 - Trabalhos relacionados
 - Nossa abordagem
- RECURSOS
- PANORAMA

Workshop de Mapeamento Visual de Coleções de Documentos

Visualização e Mineração

- DADOS VOLUMOSOS
 - Em geral representados como vetores
- VISUALIZAÇÃO
 - Transformação em gráficos e Interação com o Usuário
- MINERAÇÃO DE DADOS
 - Interpretação automática e extração de modelos
- MINERAÇÃO VISUAL
 - Combinação das técnicas
- DIMENSÕES ASSOCIADAS À INTERPRETAÇÃO DE TEXTOS

Workshop de Mapeamento Visual de Coleções de Documentos

Textos

- MINERAÇÃO DE TEXTOS (*TEXT MINING*)
 - *Classification, Categorization, Clustering*
- INFORMAÇÕES DE INTERESSE
 - Documento, Grupos de Documento,
 - Características de Coleções (Resultados de Mineração, Temas, Bibliometria, Qualidade)
 - Similaridade de Conteúdo, Estrutura
- VISUALIZAÇÃO
 - Pixel, Linha, Curva, Região
 - Pontos no Plano, Triangulações, Superfícies
 - Grafos
- MINERAÇÃO VISUAL

Workshop de Mapeamento Visual de Coleções de Documentos 7

Visualização de Domínios do Conhecimento

Knowledge Domain Visualization (KDVis)

Mapeamento de coleções de documentos e suas informações para a forma gráfica, interativa e capaz de refletir relações de interesse, de forma a conseguir interpretar a estrutura, a dinâmica e a evolução de um domínio representado textualmente.

- Temas, citações, co-citações, co-autoria, conteúdo, data de publicação.

Workshop de Mapeamento Visual de Coleções de Documentos 8

Visual Analytics

"Visual Analytics is the science of analytical reasoning supported by the highly interactive visual interface".

"Visual analytics requires interdisciplinary science beyond traditional scientific and information visualization to include statistics, mathematics, knowledge representation, management and discovery technologies, cognitive and perceptual sciences, decision sciences".

- CALL FOR PAPERS
IEEE Transactions on Visualization and Computer Graphics (TVCG) - "Visual Analytics" Special Issue in Late 2006

Workshop de Mapeamento Visual de Coleções de Documentos 9

Visual Analytics Exemplo de Relevância

- **"National Visualization and Analytics Center"** (a partir de 2004)
- **US Department of Homeland Security**
- Livro para *download* (disponível em 07/07/2005):
Illuminating the Path (versão impressa a ser publicada pela IEEE logo)
- (foco: contra-ataque ao terrorismo)

Workshop de Mapeamento Visual de Coleções de Documentos 10

Visual Analytics Exemplo de Relevância

Workshop de Mapeamento Visual de Coleções de Documentos 11

Visual Analytics Destaques

- Característica dos dados
 - Texto (primeiro da lista)
 - Bases de Dados
 - Imagem
 - Sensor
 - Vídeo
- Escalabilidade
 - da Informação
 - Visual
 - de Software
 - do Display
 - Humana

Workshop de Mapeamento Visual de Coleções de Documentos 12

ORGANIZAÇÃO

- MOTIVAÇÃO
- INTRODUÇÃO
 - Visualização e Mineração Visual
 - Visualização de Domínios do Conhecimento
 - Visual Analytics
- TÉCNICAS
 - Trabalhos relacionados
 - Nossa abordagem
- RECURSOS
- PANORAMA

Workshop de Mapeamento Visual de Coleções de Documentos

Técnicas

- Do ponto de vista de visualização
 - Pré-processamento do texto
 - De data cleaning à representação vetorial
 - Número de elementos
 - A técnica varia conforme a escala
 - Técnicas multi-escala
 - Do ponto de vista visual
 - Do ponto de vista de complexidade computacional e desempenho
 - Dimensionalidade:
 - Milhares a centenas de milhares para uma quantidade de documentos da ordem de milhares.
- Técnicas convencionais e específicas

Workshop de Mapeamento Visual de Coleções de Documentos

Técnicas Básicas

Pré-processamento do Texto

- Termos, Filtragem de Termos, Contagem de Frequência e/ou
- Extração de Relacionamentos (co-ocorrência, co-citação, citação, co-autoria, palavras-chave)

Redução de Dimensionalidade

- PCA, MDS, LSA, SOM

Agrupamento (Clustering)

Projeções

Grafos/Redes de autores ou documentos

- Refletem prioritariamente relacionamentos
- Adicionam outros atributos

Workshop de Mapeamento Visual de Coleções de Documentos

Técnicas Básicas

FDP (Force Directed Placement)

Eficiente

Lento

Alguns incrementais, outros não

Escalabilidade

- Difícil de Obter
- Chalmers 1996, 2002, 2003, 2004

Workshop de Mapeamento Visual de Coleções de Documentos

Técnicas – Texto Individual [Miller et. al 98]

Características de um texto

- Wavelets de sinais digitais a partir de palavras
- Mudança de tema
- Mudança de estilo
- 'breakpoints'

Workshop de Mapeamento Visual de Coleções de Documentos

Técnicas Coleções SOM – Skupin e outros

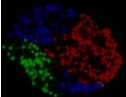
Workshop de Mapeamento Visual de Coleções de Documentos

Co-authoring Net [Ke et. al 04] InfoVis 04 Contest

? PASSOS ?

[1] Batagelj, V. & A. Mrvar *Pajek: Program Package for Large Network Analysis*. University of Ljubljana, Slovenia.

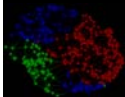
[2] Kleinberg, J.M. (2002). *Bursty and hierarchical structure in streams*. *Proceedings of the 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* pp 91-101. New York: ACM Press.



Co-authoring Net [Ke et. al 04] InfoVis 04 Contest

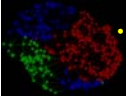
? PASSOS – Data Cleaning ?

- Identifying Major Publication Venues (called sources) (314 unique sources ==> 106 unique sources)
- Identifying Unique Keywords (1859 keywords ==> 1753 keywords)
- Identifying Unique Authors (1161 authors ==> 1036 authors)
- Citation Year (8507 ==> 8502 / 324 without year)



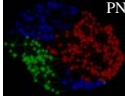
ORGANIZAÇÃO

- MOTIVAÇÃO
- INTRODUÇÃO
 - Visualização e Mineração Visual
 - Visualização de Domínios do Conhecimento
 - *Visual Analytics*
- TÉCNICAS
 - Trabalhos relacionados
 - Nossa abordagem
- RECURSOS
- PANORAMA



Recursos

- Information CyberInfrastructure (Katy Borner - Indiana)
- Citescape (Chaomei Chen - Drexel)
- IEEE InfoVis
- IV (Information Visualization Conference - Julho Londres) KDVis
- VDA (Visual Data Analysis – SPIE – Janeiro, CA (EUA))
- PNAS (**Proc. of the National Academy of Sciences**) Special Supplement em Knowledge Domain Visualization PNAS April 6, 2004 vol. 101 suppl. 1

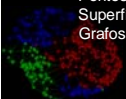


Panorama

Técnicas
um documento
vários documentos
muitos muitos documentos

Perspectivas da Visualização
documentos / grupos de documentos
autoria/co-autoria
citação
documento individual

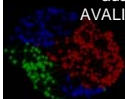
Visualização base
Pontos
Superfícies
Grafos / Redes



Panorama

Situação
Vários disponíveis – úteis mas difíceis de usar
Muitos problemas em aberto

Problemas
Muito 'data cleaning'
Quantidade de informação
Pré-filtragem absolutamente necessária
Vários níveis de informação
Desempenho dos métodos
Pré-processamento dependente de controle do usuário ou adaptação à aplicação
AVALIAÇÃO



Panorama



Nossa abordagem (até o momento)

- Displays que localizem padrões
- Exploração em nível individual e grupos
- Aproximação no display por similaridade de conteúdo
- Potencial para multi-nível
- Aproveitamento de recursos/conhecimento/potencial de SciVis
- Versões rápidas de display
- Técnicas geométricas
- Divulgação de ferramentas para exploração

