

Uma abordagem híbrida para organização flexível de documentos

Apresentação de Monografia

Nilton Vasques Carvalho Junior

Universidade Federal da Bahia

Departamento de Ciência da Computação

Orientadora: Profa. Dra. Tatiane Nogueira Rios

Contato: niltonvasques {arroba} dcc.ufba.br

2 de Junho de 2016

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de **25 terabytes/dia** (Havens et al., 2012).

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de **25 terabytes/dia** (Havens et al., 2012).
- Governos e corporações também produzem milhares de **documentos** todos os dias, tais como relatórios, formulários pesquisas de opiniões e etc.

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de **25 terabytes/dia** (Havens et al., 2012).
- Governos e corporações também produzem milhares de **documentos** todos os dias, tais como relatórios, formulários pesquisas de opiniões e etc.
- Muggleton (2006) ressalta que este cenário está além dos limites humanos para o uso e compreensão.

Introdução

- Kobayashi e Aono (2008) enfatizam que instituições estão sobrecarregadas com o processamento desse montante de dados.

Introdução

- Kobayashi e Aono (2008) enfatizam que instituições estão sobrecarregadas com o processamento desse montante de dados.
- Os dados por sua vez, estão armazenados em diversos tipos e formatos, sendo muitas vezes **não estruturados**.

Introdução

- Kobayashi e Aono (2008) enfatizam que instituições estão sobrecarregadas com o processamento desse montante de dados.
- Os dados por sua vez, estão armazenados em diversos tipos e formatos, sendo muitas vezes **não estruturados**.

Exemplos

documentos de textos, planilhas, áudios, imagens, vídeos e documentos HTML.

Introdução

As seguintes áreas vem explorando e propondo técnicas para otimizar esse processo:

- Mineração de Dados (MD)
- Aprendizado de Máquina
- Recuperação da Informação (RI)

Introdução

- Dados estruturados já possuem mecanismos eficientes de armazenamento e recuperação.
- **Documentos textuais** são recuperados através de Sistemas de Recuperação da Informação (SRI), por conta da **ausência de estruturas**.

Exemplos

Duckduckgo, Jus Brasil, IEEEExplore, ACM, Google e etc

Introdução

- Demanda crescente para desenvolvimento e aprimoramento de métodos e algoritmos que possam processar e extrair padrões dos dados.
- A extração de padrões de documentos textuais é o principal objetivo da Mineração de Textos (MT).

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 **Trabalhos relacionados**
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros

What is haplotyping and why is it important?

You hopefully know this after the previous three talks...


Conteúdo


- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
- 5 Conclusão
- 6 Trabalhos futuros**


What is haplotyping and why is it important?

You hopefully know this after the previous three talks...

Introdução

 HAVENS, T. et al. Fuzzy c-means algorithms for very large data. *IEEE Transactions on Fuzzy Systems*, v. 20, n. 6, p. 1130–1146, 2012.

 KOBAYASHI, M.; AONO, M. Vector space models for search and cluster mining. In: *Survey of Text Mining II*. Springer Science + Business Media, 2008. p. 109–127. Disponível em: <http://dx.doi.org/10.1007/978-1-84800-046-9_6>.

 MUGGLETON, S. H. 2020 computing: Exceeding human limits. *Nature*, Nature Publishing Group, v. 440, n. 7083, p. 409–410, mar 2006. Disponível em: <<http://dx.doi.org/10.1038/440409a>>.