

## Análise de projetos de Text Mining

### 1 Enunciado

Nesta atividade você deverá ler os artigos [4, 3, 2, 1, 5] e responder as questões abaixo. Uma cópia de cada artigo está disponível em <http://fbarth.net.br/cursoBigData/textMining.html>.

1. Os trabalhos [4, 3, 1, 5] fazem uso de dados no formato de texto. Descreva o pré-processamento utilizado nestes trabalhos. O que existe de comum e o que não é comum entre os trabalhos.
2. No artigo [2] na seção IV são descritos métodos baseados em conteúdo para a construção de sistemas de recomendação. O que existe de comum entre estes métodos e as técnicas utilizadas para pré-processamento dos dados nos outros artigos?
3. Os artigos [1, 3] utilizam técnicas de clustering. Qual é a justificativa para a utilização de técnicas de clustering nestes trabalhos? Estes trabalhos não utilizam clustering plano. Que outro tipo de clustering é utilizado neste trabalho? Qual é a justificativa para isto?
4. Os trabalhos [5, 4] implementam modelos preditivos. Qual é o objetivo destes modelos preditivos?
5. Na seção 4.1 do artigo [4] é descrito como os atributos textuais são pré-processados. Por favor, summarize o pré-processamento aplicado neste trabalho aos atributos textuais.

### 2 Data de entrega

Estas questões devem ser respondidas individualmente e entregues por email para [fabricao.barth@gmail.com](mailto:fabricao.barth@gmail.com) até o dia 28 de novembro de 2015.

### Referências

- [1] Fabrício J. Barth, Maria Cristina Belderrain, Nádia Quadros, Luciane L. Ferreira, and Antonio Pedro Timoszczuk. Recuperação e mineração de informações para a área criminal. In *Anais do VI Encontro Nacional de Inteligência Artificial*, 2007.
- [2] Fabrício J. Barth. User profile modeling for building recommendation systems: a theoretical study and state of the art. *Revista de Sistemas de Informação da FSMA*, 6:59–71, 2010.
- [3] Fabrício J. Barth. Using social network analysis and hierarchical clustering to identify groups in wikis. In IEEE, editor, *Collaborative Systems II - Simposio Brasileiro de Sistemas Colaborativos (SBSC-II), 2010 Brazilian Symposium of Colaborative Systems*, volume II, pages 8–11, 2010.
- [4] Helen Costa, Luiz H.C. Merschmann, Fabrício Barth, and Fabrício Benevenuto. Pollution, bad-mouthing, and local marketing: The underground of location-based social networks. *Information Sciences*, 279:123 – 137, 2014.
- [5] João Carlos Medau, Maria Cristina Belderrain, and Fabrício J. Barth. Reordenação de resultados de busca na web conforme critério de relevância definido pelo usuário. In *Anais do XI Simpósio Brasileiro de Sistemas Multimídia e Web - WebMedia*, pages 220–222, 2005.