

FCT/Unesp – Presidente Prudente
Departamento de Matemática e Computação

Visualização de Dados não Estruturados

Parte 2

Prof. Danilo Medeiros Eler
danilo.eler@unesp.br

Sumário

- Parte 1
 - Dados textuais
 - Coleções de Documentos

- Parte 2
 - Coleções de Imagens

Dados Multivariados

- Dados multivariados são aqueles que possuem mais de uma variável para cada instância dos dados

Country	GDP/capita	Public Debt	Deficit	Inflation	Unemployment
Austria	39.8	72.3	-4.6	1.7	3.9
Belgium	36.3	96.8	-4.1	2.3	6.7
Bulgaria	12.9	16.2	-3.2	3.0	11.9
Cyprus	29.0	60.8	-5.3	2.6	7.8
Czech Republic	25.0	38.5	-4.7	1.2	6.6
Denmark	36.4	43.6	-2.7	2.2	7.1
Estonia	18.5	6.6	0.1	2.7	12.8
Finland	34.9	48.4	-2.5	1.7	7.8
France	33.9	81.7	-7.0	1.7	9.9
Germany	36.1	83.2	-3.3	1.2	5.8
Greece	28.5	142.8	-10.5	4.7	16.7
Hungary	18.8	80.2	-4.2	4.7	9.9

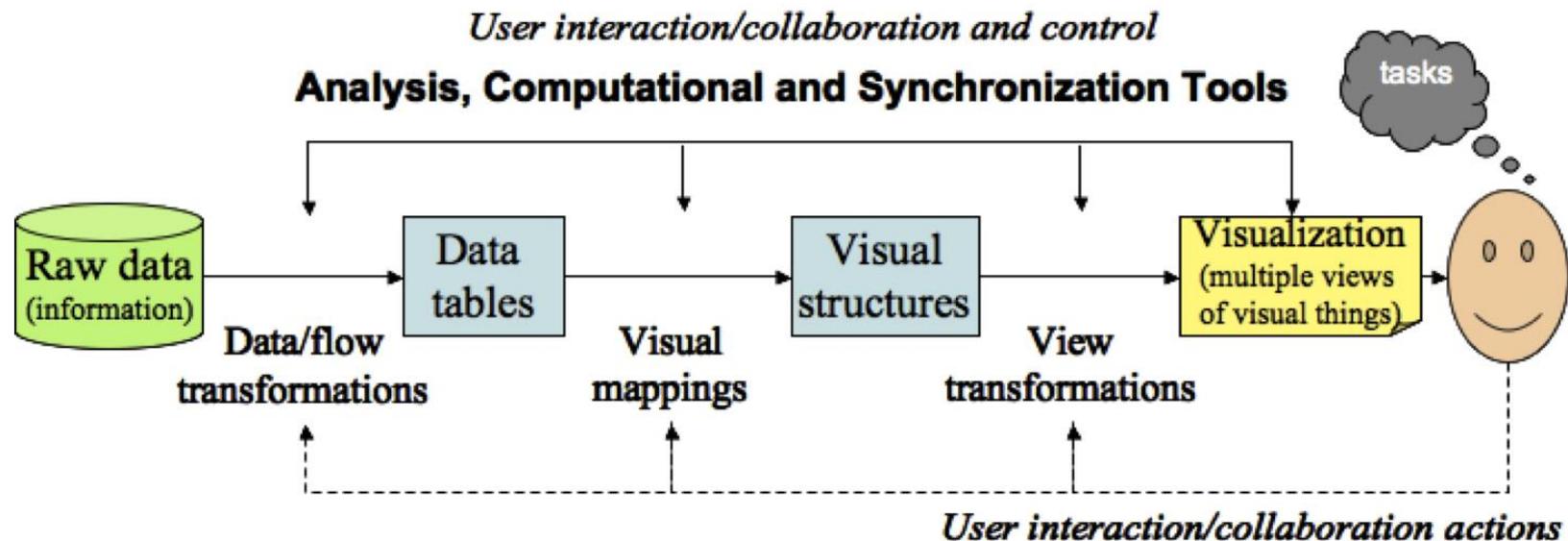
<http://www.real-statistics.com/wp-content/uploads/2013/09/eu-data.png>

Dados não Estruturados

- Alguns conjuntos de dados não possuem uma estrutura definida, por exemplo
 - Coleções de Documentos e de Imagens
- Por isso, é necessário fazer um processamento das instâncias para extrair dados para serem visualizados ou estruturar o conjunto de dados. Por exemplo,
 - Modelo de espaço vetorial para documentos
 - Espaço de características para imagens

Processo de Visualização

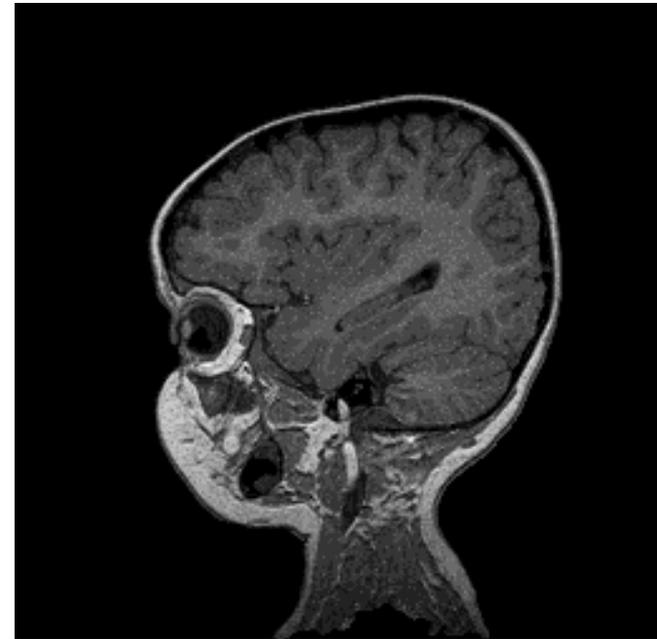
- Pipeline de visualização utilizado pela maioria dos sistemas



Visualização de Coleções de Imagens

Visualização de Coleções de Imagens

- Uma imagem pode ser vista como uma matriz de pixels, os quais podem ser divididos em três canais de cores vermelho, verde e azul



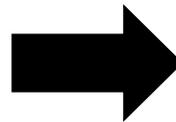
Visualização de Coleções de Imagens

- As informações contidas nas imagens não estão disponíveis e estruturadas



Visualização de Coleções de Imagens

- Assim como ocorre com as coleções de documentos, é necessário extrair dados (características) das imagens para aplicar algoritmos de mineração ou de visualização
 - Esse dados formarão os espaços de características das imagens



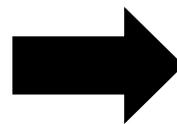
```
01. bmp;0.253651;1.103858;0.146851;1.730114;0.819273;1.228552;-0.080777;
02. bmp;0.634517;0.580094;0.138231;1.919906;0.88139;0.88948;0.054481;1.4
03. bmp;-0.208319;0.307727;0.063076;1.542225;0.194427;0.813152;0.012911;
04. bmp;-0.328327;-0.021555;-0.119767;0.74752;-0.103547;0.498694;-0.2017
05. bmp;-0.097554;0.211745;0.023506;0.981789;0.148541;1.011374;-0.064422
06. bmp;-0.062143;0.226636;0.008144;0.838526;0.193894;1.07674;-0.102697;
07. bmp;-0.333595;0.030136;-0.125789;0.296072;-0.018945;0.646808;-0.3073
08. bmp;-0.161259;0.4154;-0.016702;0.710619;0.446327;0.973397;-0.231209;
09. bmp;-0.096015;0.582253;-0.002275;0.976977;0.454398;0.882943;-0.26037
10. bmp;0.002479;0.727524;-0.037084;1.230375;0.46806;0.747805;-0.23342;0
11. bmp;-0.096521;0.45388;-0.216036;0.837342;0.105463;0.341476;-0.350149
12. bmp;0.002115;0.444415;-0.214892;0.930235;0.097837;0.297839;-0.237669
13. bmp;-0.130594;0.119448;-0.308045;0.627074;-0.144306;0.018441;-0.3633
14. bmp;0.341201;0.221144;-0.033324;1.394105;0.4916;0.509412;-0.116253;1
15. bmp;0.054547;-0.111365;-0.179589;1.003849;0.150157;0.175779;-0.22803
16. bmp;0.113357;0.230009;-0.104322;1.232718;0.52639;0.54105;-0.028232;1
```

Extração de Características

- Dado uma imagem, o método extrai características a partir dos pixels da imagem
- Na literatura existem diversos métodos para extrair características de imagens
 - Por exemplo, filtros de Gabor, matriz de co-ocorrência, momentos de Zernik, medidas estatísticas, entre outros

Extração de Características

- Por exemplo, podemos extrair a média dos canais de cores de uma imagem colorida como características para representar essa imagem

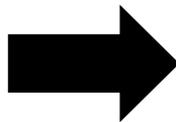
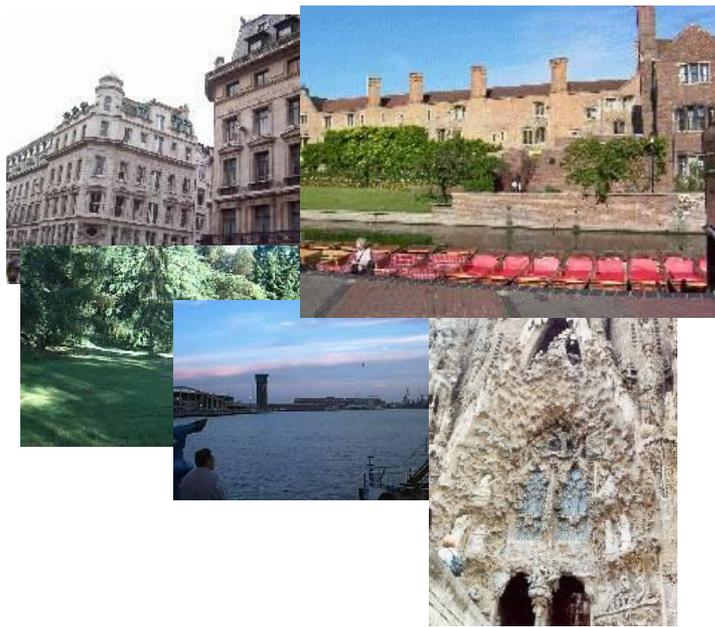


Vetor de Características
[100, 200, 150]

Espaço de Características

- O conjunto de vetores computados a partir das imagens de uma coleção forma o Espaço de Características de uma coleção de imagens

Espaço de Características



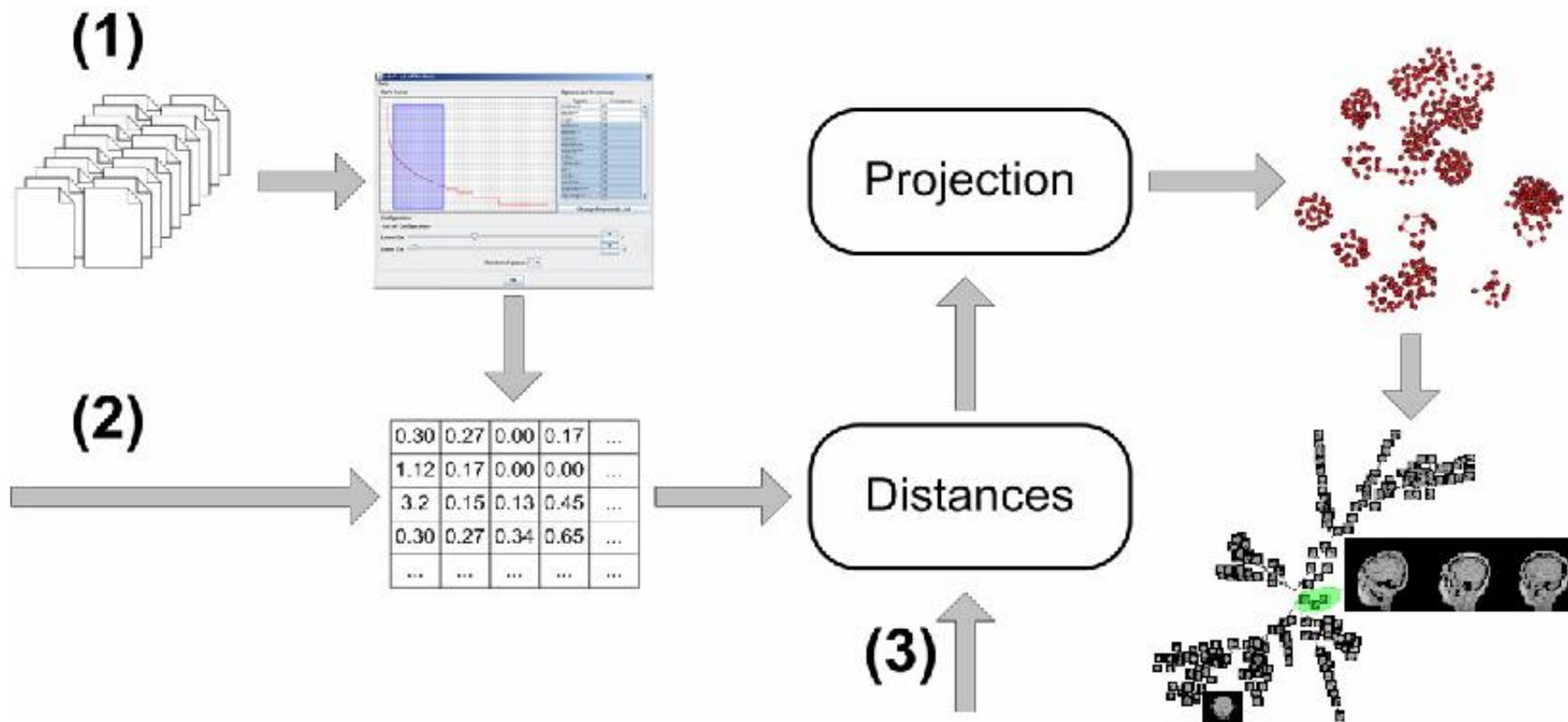
Img-ID	MediaR	MediaG	MediaB
img1	100	200	150
Img2	140	170	110
Img3	50	200	110
....

Espaço de Características

- Uma vez que o espaço de características foi construído, a coleção de imagens foi estruturada
- Assim, é possível aplicar algoritmos de mineração e de visualização
 - Por exemplo, podemos utilizar projeções multidimensionais para reduzir o espaço dimensional e visualizar as relações no plano 2D

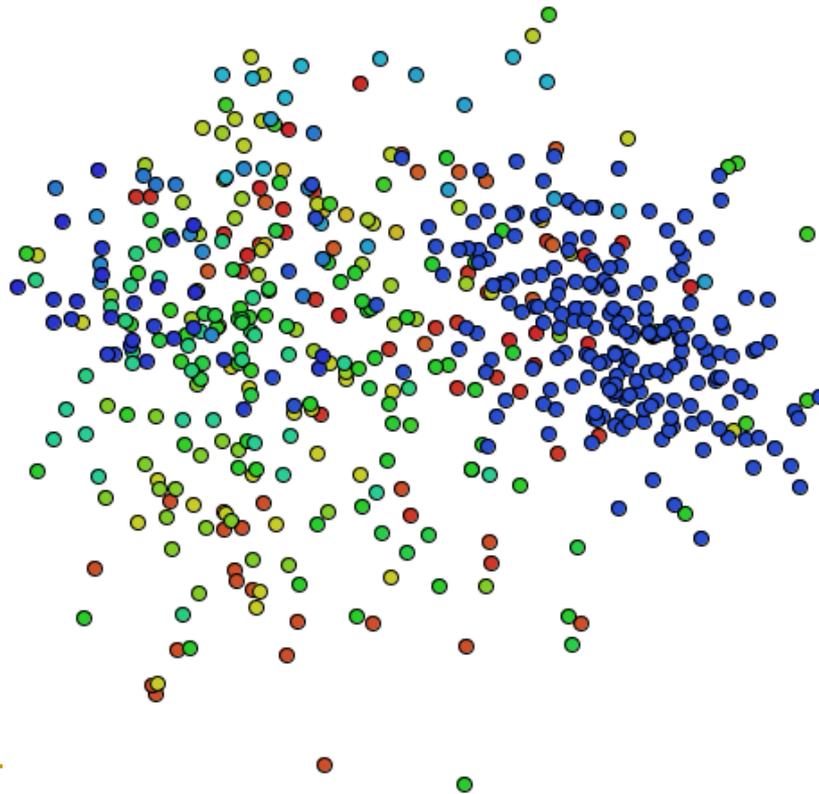
Projeções a partir de Coleções de Imagens

- Segue o mesmo pipeline apresentado para documentos



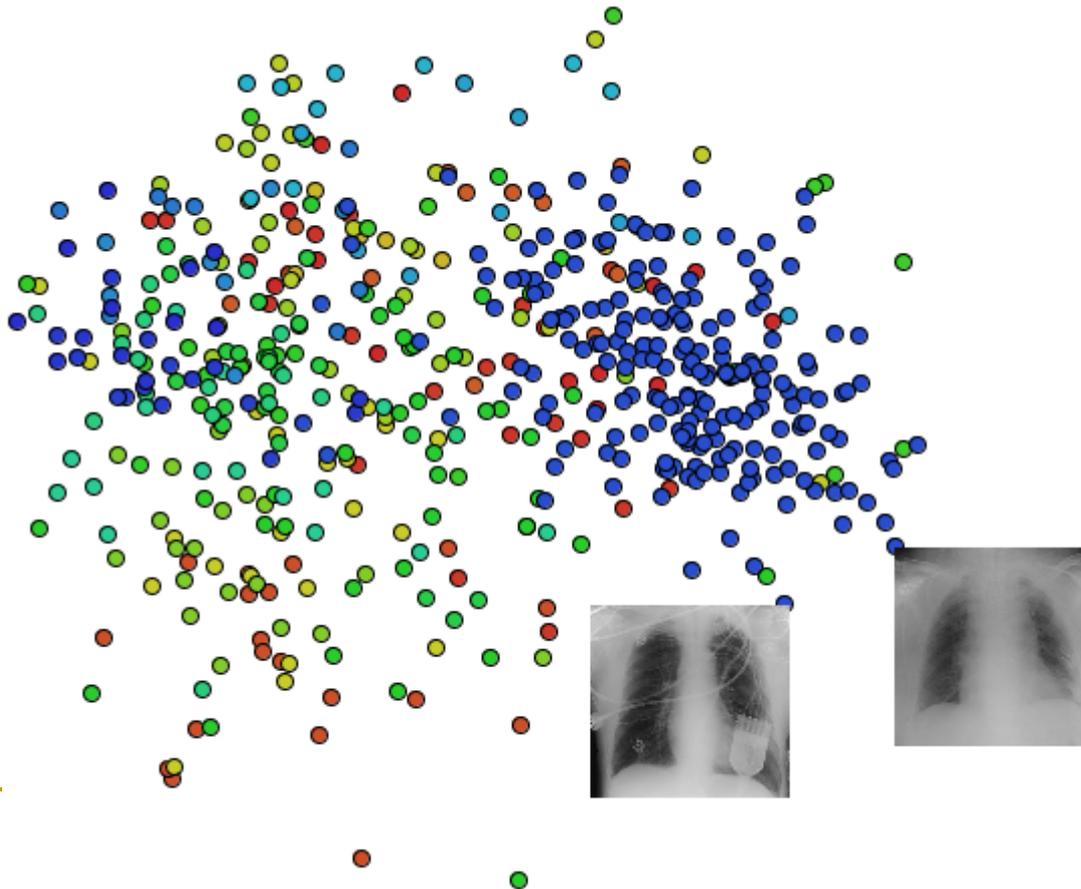
Projeções a partir de Coleções de Imagens

- Visualização de uma coleção de imagens de raios X
 - As características foram extraídas por meio de Wavelets



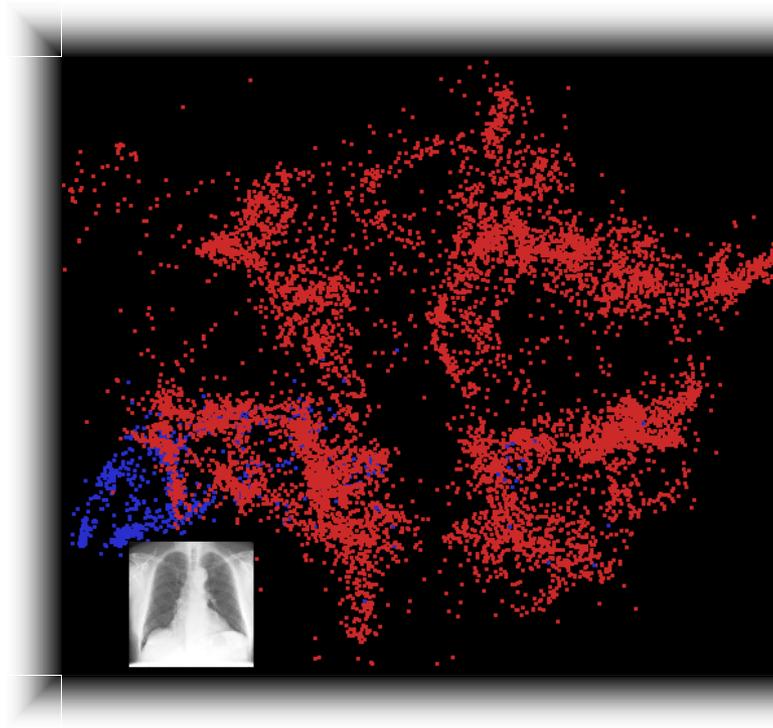
Projeções a partir de Coleções de Imagens

- Visualização de uma coleção de imagens de raios X
 - As características foram extraídas utilizando Wavelets

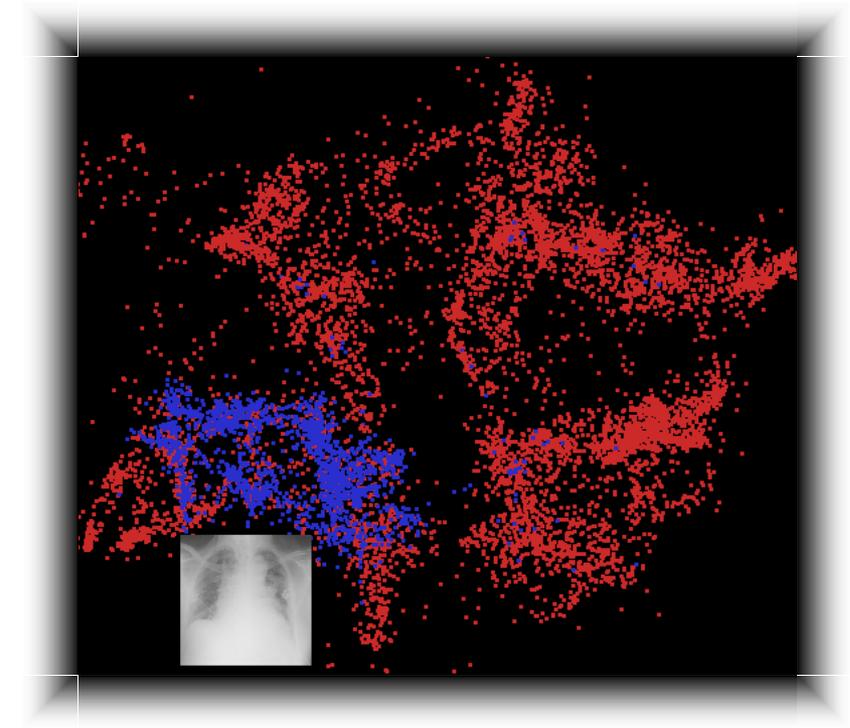


Projeções a partir de Coleções de Imagens

- Visualização de uma coleção de imagens de raios X
 - As características foram extraídas utilizando Wavelets



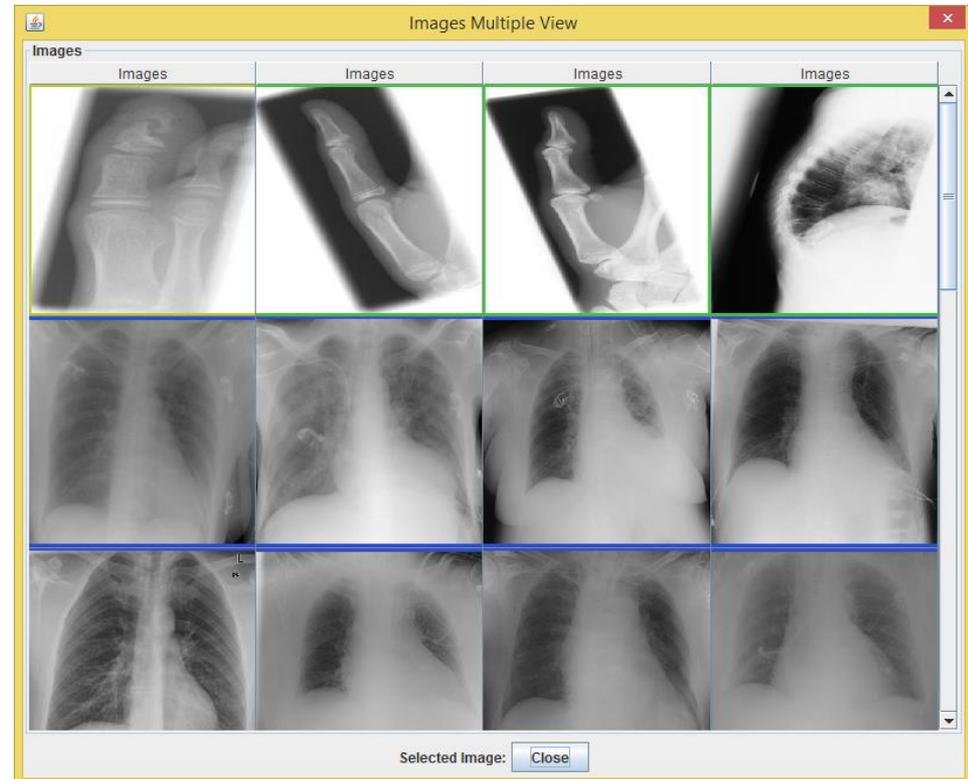
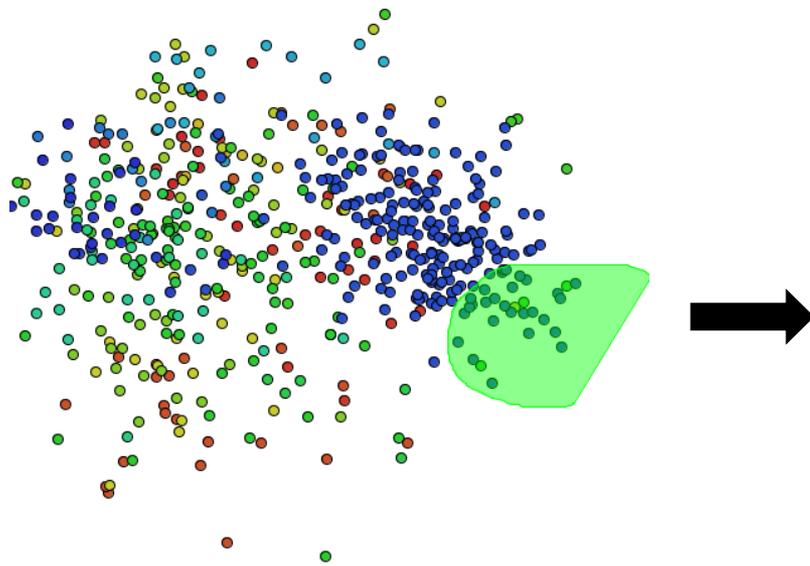
Classe 108



Classe 111

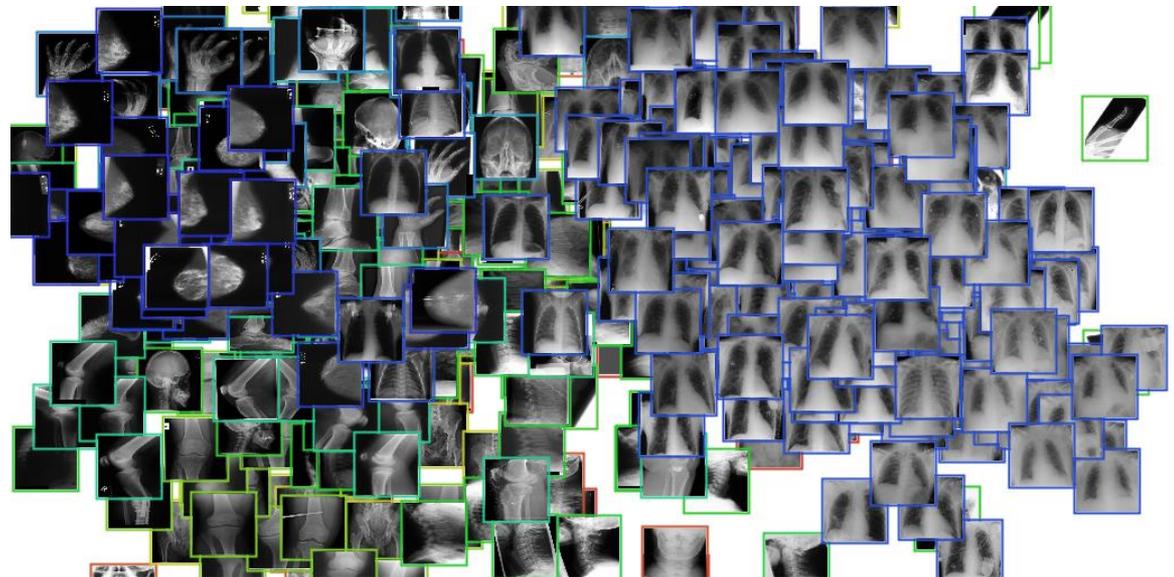
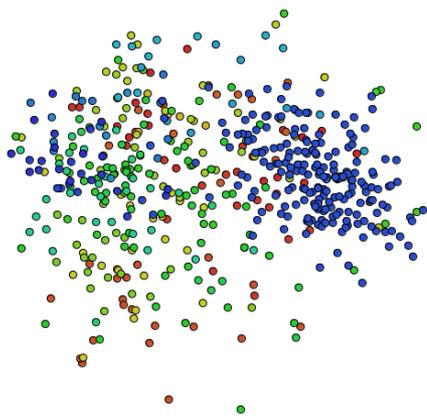
Projeções a partir de Coleções de Imagens

- Visualização de uma coleção de imagens de raios X
 - As características foram extraídas utilizando Wavelets

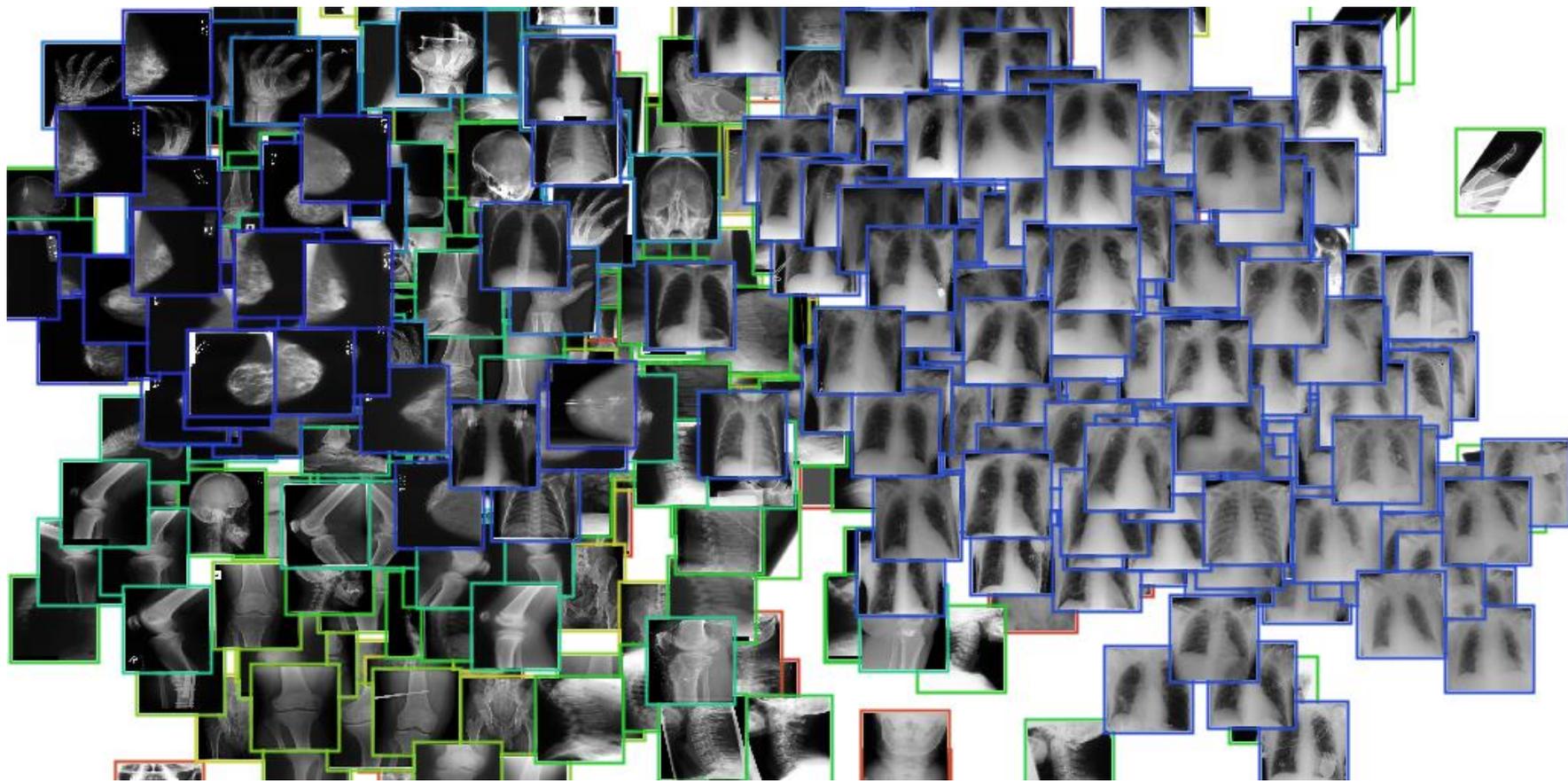


Projeções a partir de Coleções de Imagens

- Visualização de uma coleção de imagens de raios X
 - As características foram extraídas utilizando Wavelets



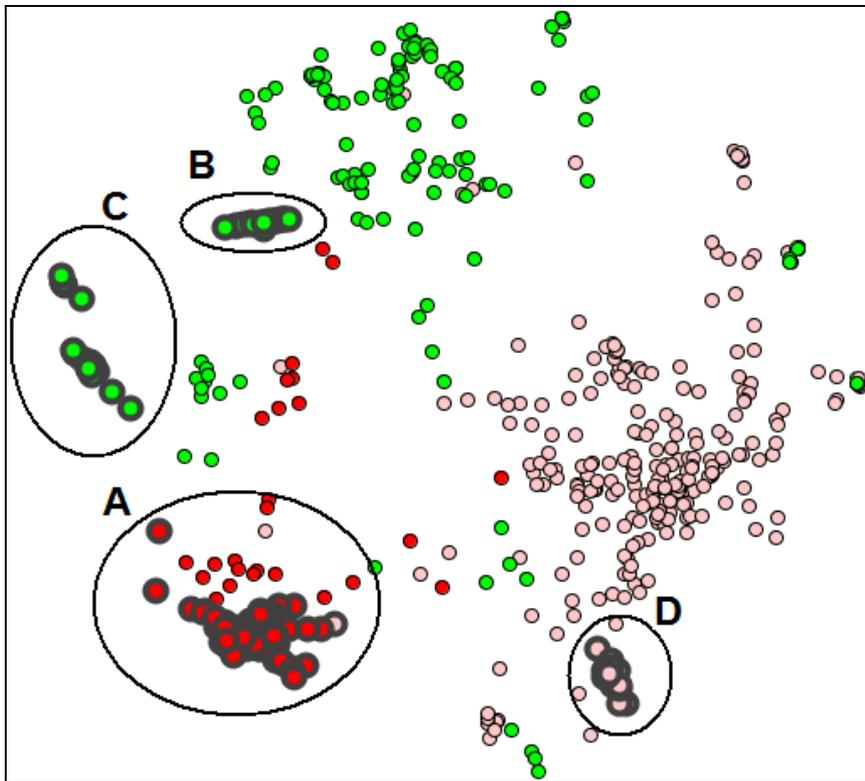
Projeções a partir de Coleções de Imagens



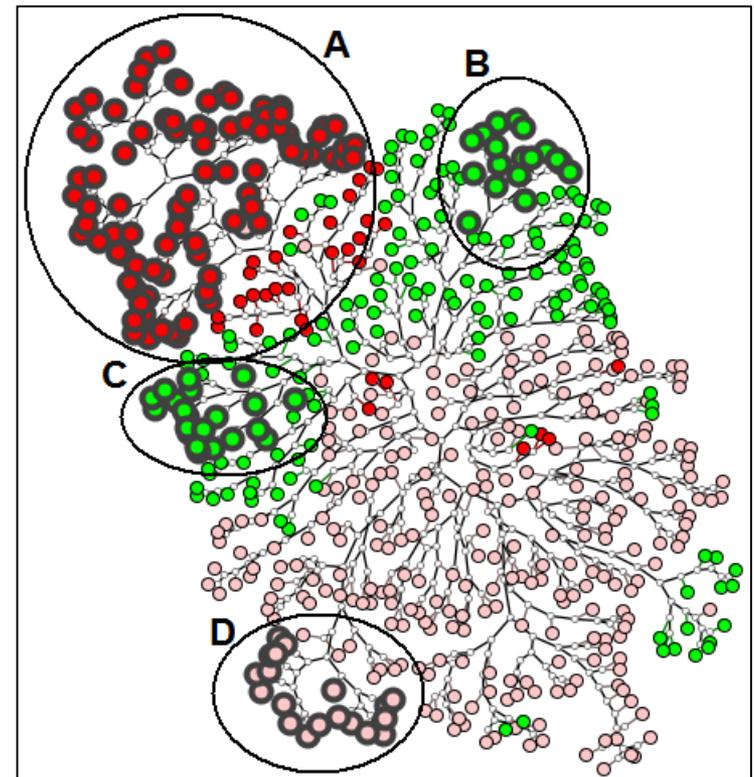
Visualização de Coleções de Imagens

- Representação baseada em árvore

Projeção

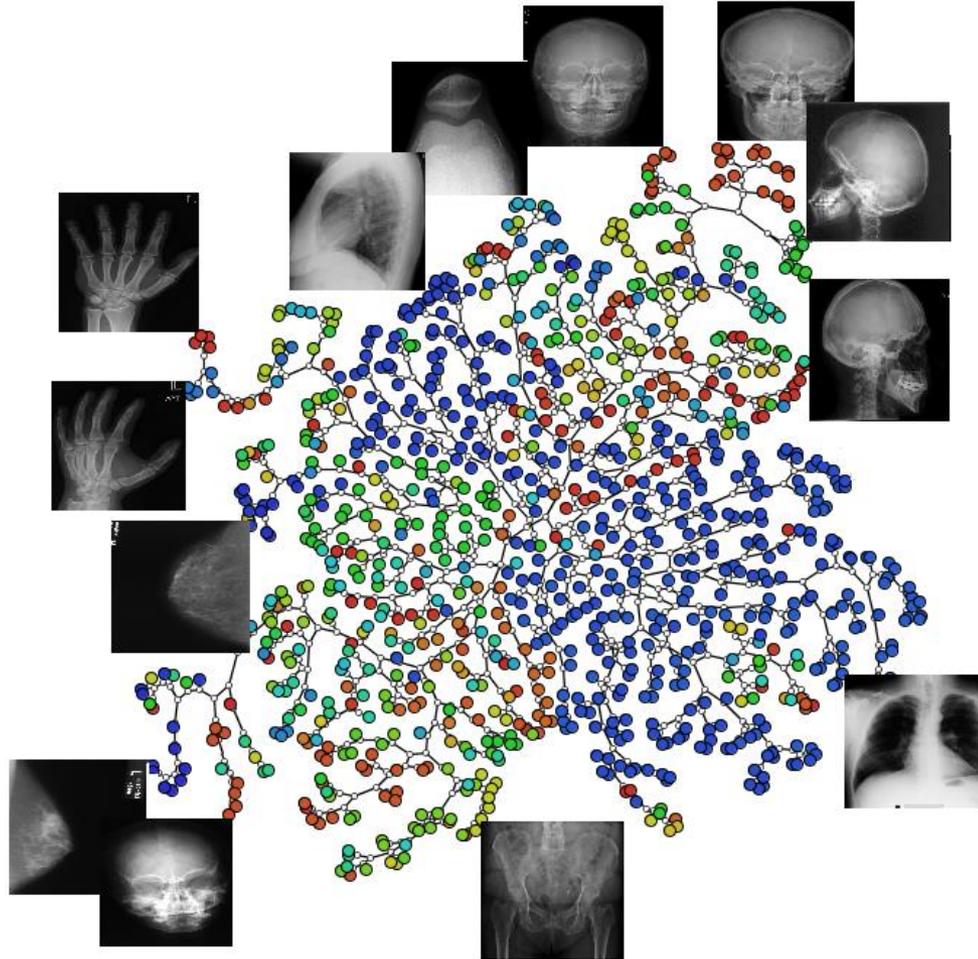


Árvore



Visualização de Coleções de Imagens

- Representação baseada em árvore

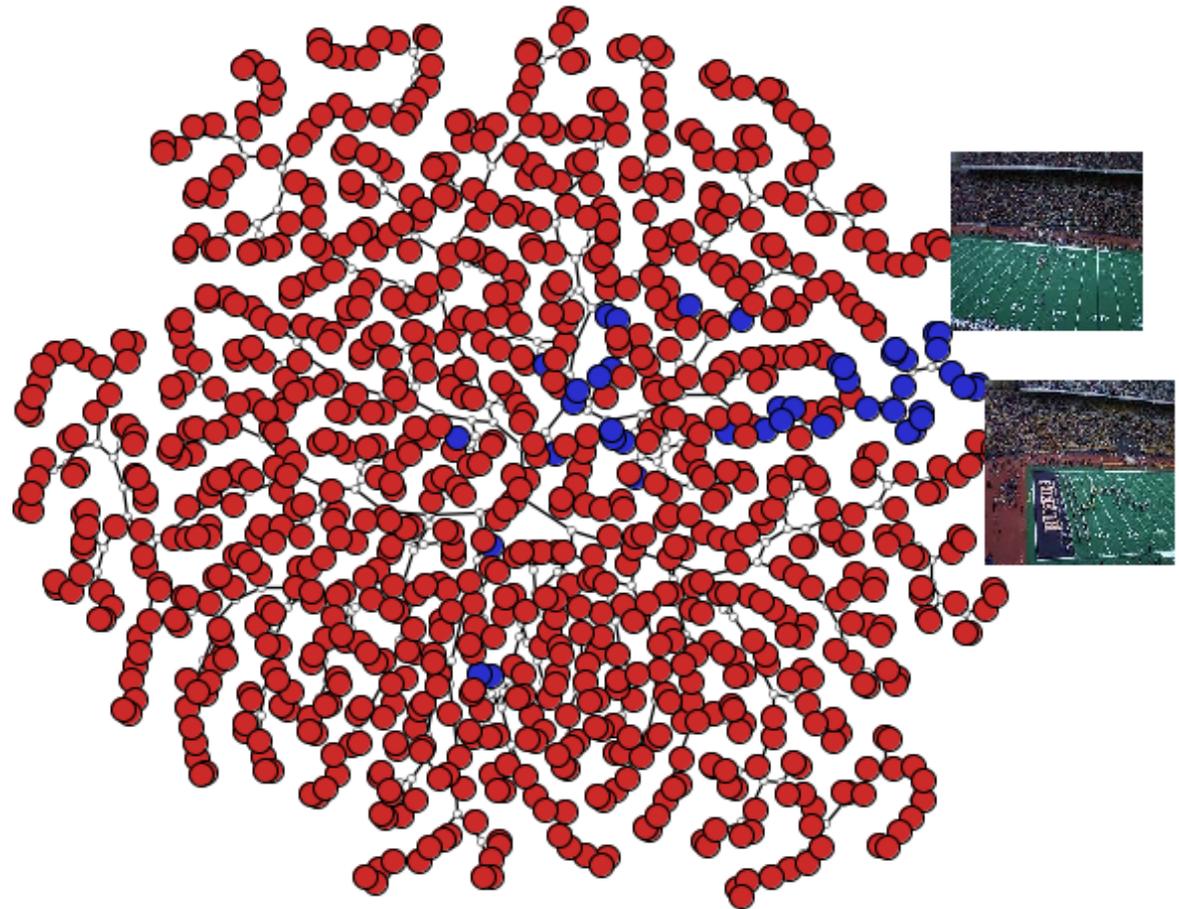


Visualização de Coleções de Imagens

- Alguns conjuntos de dados de imagens possuem informações textuais associadas
 - Nesse caso, podemos utilizar as duas representações para explorar o conjunto
 - A matriz de termos x documentos pode ser construído a partir da informação textual
 - O Espaço de Características pode ser construído a partir das imagens

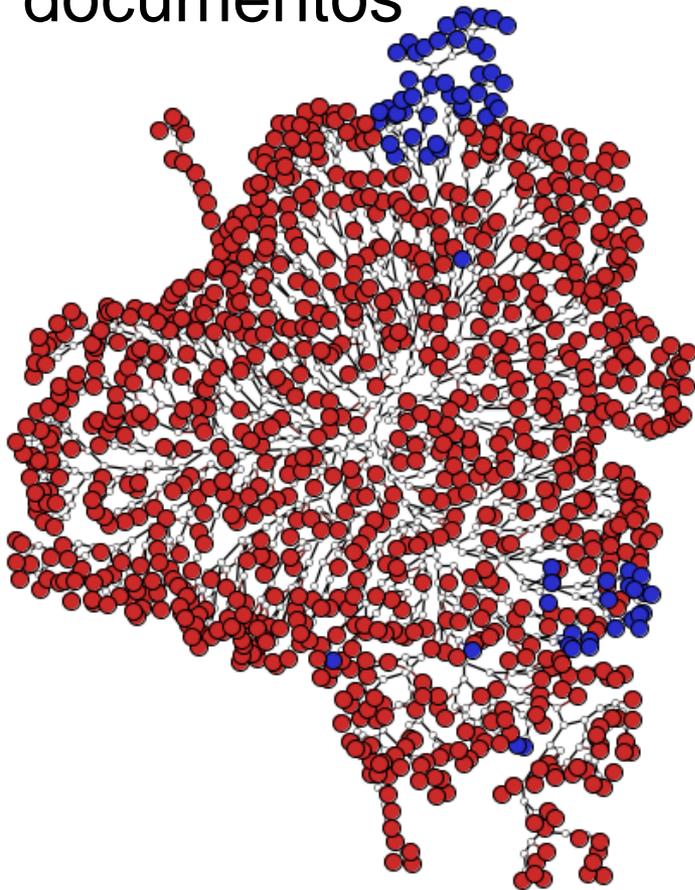
Visualização de Coleções de Imagens

- Palavra chave
 - “stadium”
- Árvore construída a partir do espaço de características



Visualização de Coleções de Imagens

- Palavras chave: “stadium” e “temple”
- Árvores construídas a partir da matriz de termos x documentos



Visualização de Coleções de Imagens

- Zoom in na seleção



Referências

- Ward, M., Grinstein, G. G., Keim, D.
 - Interactive data visualization foundations, techniques, and applications. Natick, Mass., A K Peters, 2a Edição, 2010.
 - Capítulo 10 (Text and Document Visualization)
- G. Salton, A. Wong, and C. S. Yang.
 - “A Vector Space Model for Automatic Indexing.” Commun. ACM 18:11 (1975), 613–620
- M. Wattenberg and F. B. Viégas
 - “The Word Tree, an Interactive Visual Concordance.” IEEE Transactions on Visualization and Computer Graphics 14:6 (2008), 1221–1228.
- Jonathan Feinberg.
 - “Wordle Home Page.” <http://www.wordle.net/>, accessed August 31, 2009.
- WordTree
 - IBM. “Many Eyes Home Page.” <http://manyeyes.alphaworks.ibm.com/>, accessed August 31, 2009.

Referências

- W. B. Paley.
 - “TextArc: Showing Word Frequency and Distribution in Text.” Poster presented at IEEE Symposium on Information Visualization, Boston, MA, October 27–November 1, 2002.
 - <http://www.textarc.org/>
- D. A. Keim and D. Oelke.
 - “Literature Fingerprinting: A New Method for Visual Literary Analysis.” In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST 2007), pp. 115–122. Los Alamitos, CA: IEEE Computer Society Press, 2007.
- T. Kohonen.
 - Self-Organizing Maps, Springer Series in Information Sciences, 30, Third edition. Berlin: Springer, 2001.
- Steffen Lohmann ; Florian Heimerl ; Fabian Bopp ; Michael Burch ; Thomas Ertl
 - Concentri Cloud: Word Cloud Visualization for Multiple Text Documents. In 19th International Conference on Information Visualisation, 2015
 - <https://ieeexplore.ieee.org/abstract/document/7272588>

Referências

■ Projection Explorer (PEx)

- F. V. Paulovich, M. C. F. Oliveira, and R. Minghim
 - “The projection explorer: A flexible tool for projection-based multidimensional visualization”, in XX Brazilian Symposium on Computer Graphics and Image Processing. Washington, DC, USA: IEEE Computer Society, 2007, pp. 27–36.
 - <http://vis.icmc.usp.br/vicg/tool/1/projection-explorer-pex>

■ Projection Explorer for Images (PEx-Image)

- D. M. Eler, M. Nakazaki, F. Paulovich, D. Santos, G. Andery, M. Oliveira, J. E. S. Batista, and R. Minghim
 - “Visual analysis of image collections”, The Visual Computer, vol. 25, no. 10, pp. 923–937, 2009.
 - <https://github.com/daniloeler/PEx-Image>