Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Genome comparison based on raw read and paired read data.

Diploma thesis

2016
Vladimír Macko

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Genome comparison based on raw read and paired read data.

Diploma thesis

Študijný program:	Informatika
Študijný odbor:	2508 Informatika
Školiace pracovisko:	Katedra informatiky
Školiteľ:	Mgr. Tomáš Vinař, PhD

Bratislava, 2016
Vladimír Macko

Pod'akovanie:

Abstrakt

Kľúčové slová:

Abstract

Keywords:

Obsah

Zoznam obrázkov

Zoznam tabuliek

Introduction

Kapitola 1

Something else

1.1 section

Kapitola 2

Final

2.1 Subfinal

2.1.1 sub sub final

Literatúra

[1] Databáza patentov. http://patents.reedtech.com/pgrbft.php. (navštívené dňa 20.4.2016).

[2] Databáza vedeckých článkov zo životných vied. http://www.ncbi.nlm.nih.gov/pubmed. (navštívené dňa 15.1.2016).

[3] Googlom zverejnená databáza patentov. https://www.google.com/googlebooks/uspto-patents.html. (navštívené dňa 15.1.2016).

[4] Zoznam desiatich najprofitujúcejších priemyslov v usa za rok 2014 podľa forbes. http://www.forbes.com/sites/liyanchen/2015/09/23/the-most-profitable-industries-in-2015/#6410eb587712. (navštívené dňa 2.3.2016).

[5] Nástroj od googlu na vyhľadávanie súvisiacich publikácií k patentom. https://www.google.com/patents/related, 2016. (navštívené dňa 27.1.2016).

[6] Akiko Aizawa. An information-theoretic perspective of tfidf measures. *Information Processing & Management*, 39(1):45–65, 2003.

[7] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, June 2010. Oral Presentation.

[8] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. Ö'Reilly Media, Inc.", 2009.

[9] BSR tím. Know your field of research! www.scicurve.com, 2014. (navštívené dňa 28.1.2016).

[10] Aaron M Cohen and William R Hersh. A survey of current work in biomedical text mining. *Briefings in bioinformatics*, 6(1):57–71, 2005.

[11] Hal Daumé III and Daniel Marcu. Induction of word and phrase alignments for automatic document summarization. *Computational Linguistics*, 31(4):505–530, 2005.

[12] Atsushi Fujii. Enhancing patent retrieval by citation analysis. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 793–794. ACM, 2007.

[13] Kayvan Kousha and Mike Thelwall. Patent citation analysis with google. *Journal of the Association for Information Science and Technology*, 2015.

[14] Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284, 1998.

[15] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. *arXiv preprint arXiv:1405.4053*, 2014.

[16] Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, et al. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.

[17] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[18] Fernando Pérez and Brian E. Granger. IPython: a system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

[19] PostgreSQL Global Development Group. PostgreSQL. `http://www.postgresql.org`, 2016. (navštívené dňa 23.1.2016).

[20] PostgreSQL Global Development Group. Úvod do práce s tsv vektormi a dátovými typmi. `http://www.postgresql.org/docs/8.3/static/datatype-textsearch.html`, 2016. (navštívené dňa 18.1.2016).

[21] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[22] USPTO. Návod na používanie portálu USPTO. `http://www.uspto.gov/sites/default/files/documents/7_Step_US_Patent_Search_Strategy_Guide_2015_rev.pdf`, 2015. (navštívené dňa 22.1.2016).

[23] Shuo Xu, Lijun Zhu, Xiaodong Qiao, Qingwei Shi, and Jie Gui. Topic linkages between papers and patents. In *Proceedings of the 4th International Conference on Advanced Science and Technology*, pages 176–183. Science and Engineering Research Support soCiety (SERSC), 2012.

[24] Xiaobing Xue and W Bruce Croft. Automatic query generation for patent search. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 2037–2040. ACM, 2009.

[25] Mo Yu and Mark Dredze. Learning composition models for phrase embeddings. *Transactions of the Association for Computational Linguistics*, 3:227–242, 2015.

Appendix A

TODO