Comparing Patent Similarity to Detect Potential Competitors

Saurabh Jaju, Jeffrey Hsu, and Cameron Bell

The Problem and Why it is Interesting

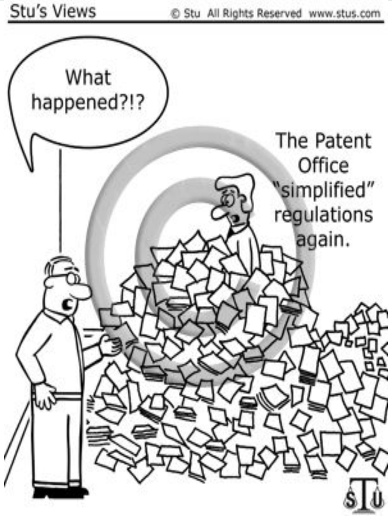
OUR MISSION
PRETTY MUCH WHATEVER OUR COMPETITION DOES, BUT SIX MONTHS LATER
TOM FISH BURNE

© marketoonist.com

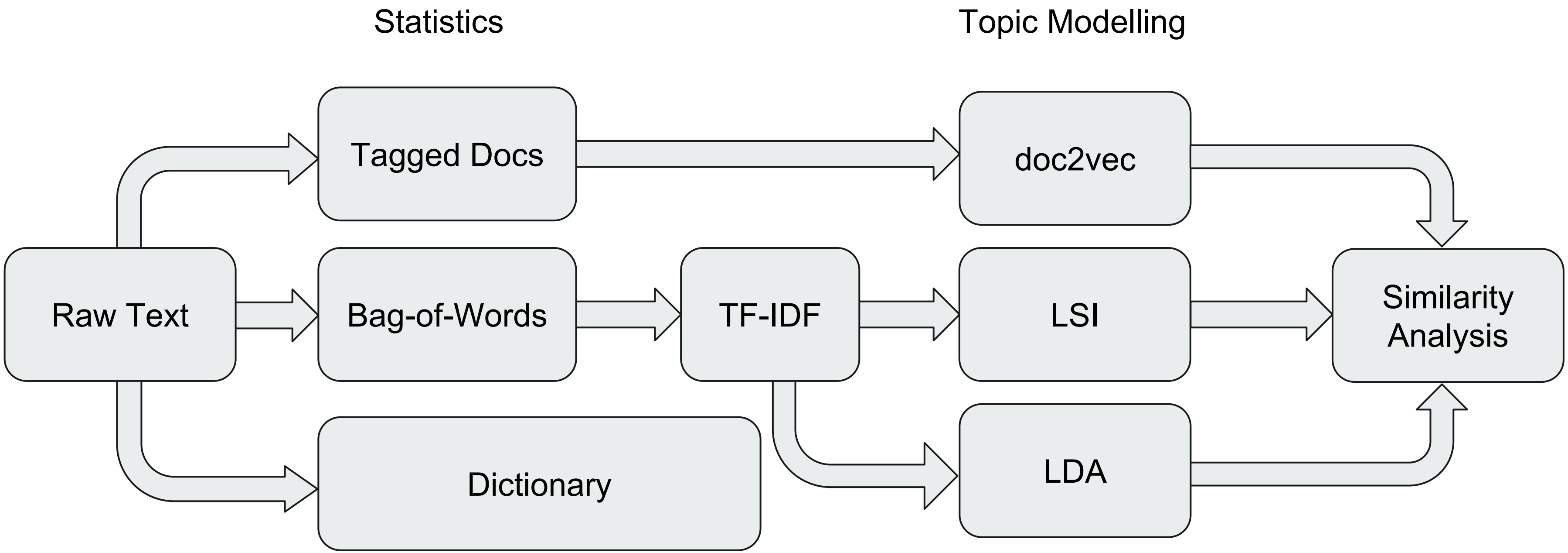
- It is not always straightforward to identify (potential) competitors
- Competitor analysis heavily affects business strategy
- It is a major factor in Porter's five force model for analysing the profitability of the industry over time
- vs IBM in smartphone market? Is it possible?

The Data

- The data is collected from US Patent office 1970 onwards.
- The fulltext data is available in CSV, XML, and other formats.
- Each patent file contains details like : applicant information, abstract, claims, text describing diagrams, and references.


Stu's Views
© Stu All Rights Reserved www.stus.com
What happened?!?
The Patent Office "simplified" regulations again.
STU

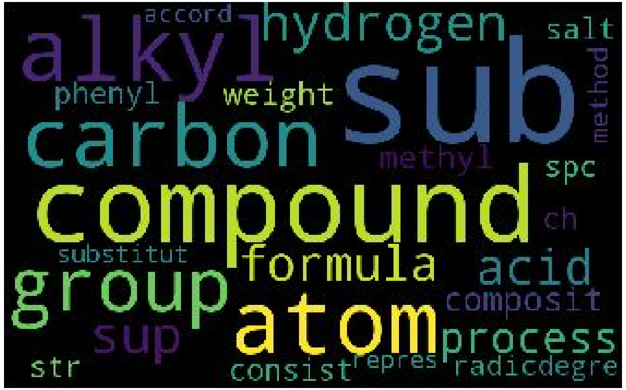
Analysis Architecture

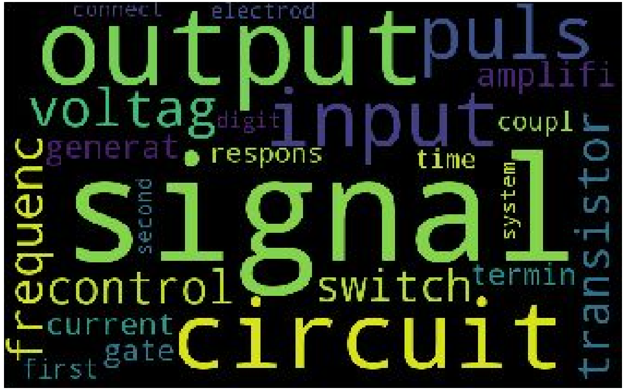


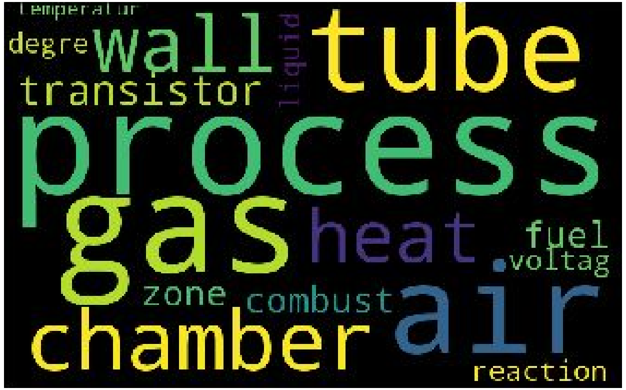
Design Choices

- To allow for scaling up, everything is read (and stored) line by line instead of loaded into memory
- Able to update the models with new patents
- Materialize models on disk

Results


accord
alkyl
hydrogen
salt
sub
phenyl
weight
method
carbon
methyl
spc
compound
ch
substitut
formula
acid
group
composit
sup
atom
process
str
consist
repres
radicdegre


connect
electrod
output
puls
amplifi
voltag
digit
input
coupl
generat
respons
time
frequenc
second
signal
system
transistor
control
switch
termin
current
circuit
gate
first

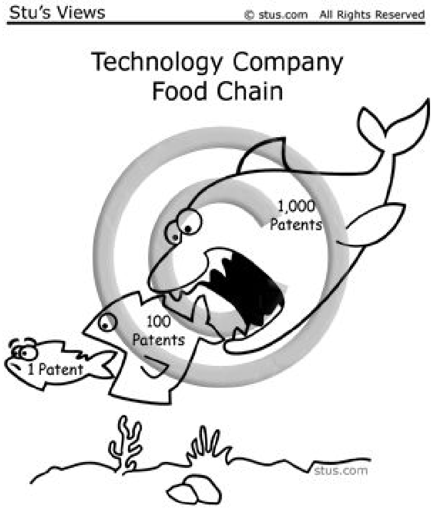

temperatur
degre
wall
liquid
tube
transistor
process
gas
heat
fuel
voltag
zone
combust
air
chamber
reaction

Results

- Demo
- Visualization

Future Scope

- Compare companies based on their patents
- Full 7-million document corpus
- Compare patents across and within industries
- Use timelines of similar patent filings and eventual entry into the industry to develop predictive models
- Use patent tags to evaluate accuracy of different semantic models



Conclusion

- Developed a simple, full text patent similarity comparison system using the following workflow:

  Raw Text → Dictionary → Bag of words → tf-idf→ LSA / LDA

  → Doc2Vec

- Check out our **Github repo**

References

[1] Measures for textual patent similarities: a guided way to select appropriate approaches. Martin G. Moehrle Scientometrics (2010) 85:95–109 DOI 10.1007/s11192-010-0243-3

[2] Similarity Analysis of Patent Claims Using Natural Language Processing Techniques .Kishore Varma Indukuri, Anurag Anil Ambekar, Ashish Sureka International Conference on Computational Intelligence and Multimedia Applications 2007

[3] Towards content-oriented patent document processing. Leo Wanner et. al World Patent Information 30 (2008) 21–33

[4] Patent-to-Patent Similarity: A Vector Space Model. KA Younge SSRN Papers

[5] Software Framework for Topic Modelling with Large Corpora https://github.com/RaRe-Technologies/gensim Rehurek, Radim, and Sojka, Petr. 2010

[6] Word Cloud: A little word cloud generator in Python. https://github.com/amueller/word_cloud Mueller, Andreas. 2017

[7] pyLDAvis: Python library for interactive topic model visualization. https://github.com/bmabey/pyLDAvis Mabey, Ben 2014