

Search Engines

Exercise 4: Text Processing

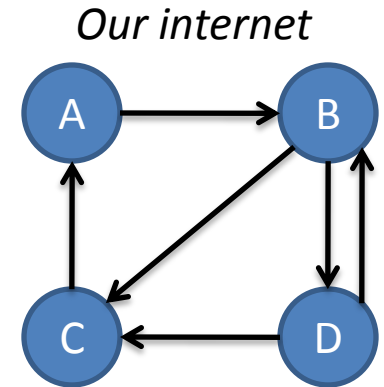
Dustin Lange & Saeedeh Momtazi

19 May 2011



Task 1: PageRank

- Apply the PageRank algorithm to *our internet*.
- Iteration 0: PageRank is equally distributed before the first iteration:
 $PR(A)=PR(B)=PR(C)=PR(D)=1/4$
- With a random visit probability of $\lambda = 0.2$, compute the first four iterations and always round to three places after the comma. Show your calculations in detail for the first two iterations. Do the values converge?





Task 2: Stemming

- Select a German and an English text (from the texts you chose in the last exercise). Process both texts using two stemmers: The Porter stemmer and the German stemmer from Snowball (<http://snowball.tartarus.org/>).
- For each text, find 5 examples of differences in the stemming (with the different stemmers) that could have an impact on ranking.
- Give example queries for which you would expect different results. (Optional)



Task 3: Text Processing

- Text processing techniques have an important impact on information retrieval.
- Their contribution is not always apparent (improving the IR performance).
- They provide benefits that go beyond quantitative metrics to support broader tasks and richer experiences.
- This is a brainstorming task to find the role of text processing techniques in search engines.



Task 3: Text Processing

For each of the following techniques, give an example, a new idea, or a new challenge which shows how the technique can contribute in information retrieval:

- Language Modeling (N-grams)
- Machine Translation
- Named Entity Recognition
- Parsing
- Part of Speech Tagging
- Relationship Extraction
- Stemming
- Spell Correction
- Summarization
- Text Categorization
- Word Prediction
- Word Sense Disambiguation



Task 3: Text Processing

Example for Language Modeling (N-grams):

new companies in York

companies in New York



new companies in York X Search

About 280,000,000 results (0.22 seconds) Advanced search

- Everything
- Images
- Videos
- News
- Shopping
- Realtime
- More

- Any time
- Latest
 - Past 24 hours
 - Past week
 - Past month
 - Past year
 - Custom range...
- More search tools

[The Press, York - news, sports, leisure, houses, jobs, cars ...](#)
 18 May 2011 ... York news and York sport coverage from The Press newspaper in York. ... "distant" says steel company - Firms urged to back York 10K race ... Looking for a new career? Find a job in York and all around North Yorkshire ...
 News - York City FC - Jobs - Announcements
www.yorkpress.co.uk/ - Cached - Similar

[Denise Stuart, of York Enterprise, is tasked with luring new ...](#)
 2 Nov 2010 ... CALL Denise Stuart a human bridge over troubled waters, and ...
www.yorkpress.co.uk/news/.../news/8489267.Denise_Stuart_is_tasked_with_luring_new_businesses_to_York/ - Cached
[Show more results from yorkpress.co.uk](#)

[York Water Company News - The New York Times](#)
 York Water Company financial and business news, updates, and information from The New York Times and other leading providers.
topics.nytimes.com/Business/Companies/ - Cached - Similar

News for new companies in York

 [MarketSharing, a New York-based company, latest iteration of ...](#)
 6 hours ago
 Amato said he got the idea for MarketSharing while running Show Media, a New York company that puts billboards on top of taxicabs. ...
[New York Daily News - 7 related articles](#)
[PRESS DIGEST - New York Times business news - May 18](#)
[Reuters - 28 related articles - Shared by 10+](#)

[Seacoast Online - Portsmouth NH, York ME, Hampton NH, Exeter NH ...](#)
 While I was in New York City last week, I went through the street where all the Korean restaurants are and spied a Korean Fried Chicken joint. ...
www.seacoastonline.com/ - Cached - Similar

[By any measure, New York's economy depends heavily on small businesses](#)
 File Format: PDF/Adobe Acrobat - Quick View
 of all firms in New York State. There were also over 1.5 million firms without ... To date, \$449 million has been invested in 145 New York companies. ...
www.osc.state.ny.us/reports/other/smallbusinessreport030410.pdf - Similar

[My Account - New York & Company](#)
 Nyandcompany.com - Find trendy, sexy, comfortable, affordable clothing online. Womens apparel, accessories and footwear. Clothing available in sizes 2-18, ...
www.nyandcompany.com/.../login_redirect.jsp?_ - Cached - Similar

[List of Motor Vehicle Dealers \(New and Used\) Companies in North ...](#)
 We have 36 Motor Vehicle Dealers (New and Used) Companies in North York, Canada. See contact info and background details for free.
www.manta.com/.../York/motor_vehicle_dealers_new_and_used_-B11FF/ - Cached

[Category: Companies based in New York City - Wikipedia, the free ...](#)
 List of companies based in New York City, New York, United States. See also: List of Architecture Firms based in New York City, New York, United States.



companies in New York Search

About 506,000,000 results (0.16 seconds) Advanced search

- Everything
- Images
- Videos
- News
- Shopping
- Places
- Realtime
- More

- Any time
- Latest
 - Past 24 hours
 - Past week
 - Past 3 weeks
 - Past month
 - Past year
 - Custom range...
- More search tools

Places for Companies near New York, NY

- A** American Express Co - 15 reviews - Place page
www.americanexpress.com - 200 Vesey St, New York - (212) 640-2000
- B** New York Life Insurance Co - 1 review - Place page
newyorklife.com - 51 Madison Avenue, New York - (212) 576-7000
- C** Brown Brothers Harriman & Co - 7 reviews - Place page
www.bbh.com - 140 Broadway, New York - (212) 483-1818
- D** McKinsey & Company Inc - 30 reviews - Place page
www.mckinsey.com - 485 Madison Avenue # 10, New York - (212) 446-7000
- E** Battery Dance Company - 3 reviews - Place page
www.batterydanceco.com - 380 Broadway # 5, New York - (212) 219-3910
- F** Paul Taylor Dance Company - 15 reviews - Place page
www.ptdc.org - 551 Grand Street, New York - (212) 431-5562
- G** Atlantic Theater Company - 58 reviews - Place page
www.atlantictheater.org - New York - (212) 645-8015

More results near New York, NY »

Companies near: Brooklyn - Manhattan - Water - Eighth Avenue - 7th Avenue

List of New York companies - Wikipedia, the free encyclopedia
List of New York companies includes notable companies that are, or once were, headquartered in New York. Contents. 1 0-9; 2 A; 3 B; 4 C; 5 D; 6 E; 7 F; 8 G ... 0-9 - A - B - C
en.wikipedia.org/wiki/List_of_New_York_companies - Cached - Similar

List of corporations based in New York City - Wikipedia, the free ...
Un sourced material may be challenged and removed. (July 2009). This is a ...
en.wikipedia.org/_/List_of_corporations_based_in_New_York_City - Cached - Similar

Category:Companies based in New York City - Wikipedia, the free ...
Pages in category "Companies based in New York City". The following 200 ...
en.wikipedia.org/_/Category:Companies_based_in_New_York_City - Cached - Similar
Show more results from wikipedia.org

News for Companies in New York

-  MarketSharing, a New York-based company, latest iteration of...
6 hours ago
MarketSharing, a New York-based company that's launching Wednesday, aims to do just that - with a twist. Founded by John Amato, a 30-year-old entrepreneur ...
New York Daily News - 7 related articles
- Malaysia To Tie Up With US Companies For Oil Palm Biomass Based ...
Bernama - 130 related articles



Submissions & Next Exercise

- Submissions:
 - Create slides to present your solution.
 - Send us your presentation
 - as PDF or PPT(X) or ODP:
SearchEngines4[Name1][Name2].[pdf|ppt|pptx|odp]
 - via e-mail with subject: *Search Engines 4*
 - to *dustin (dot) lange (at) hpi (dot) ...*
 - until **08 June 2011, 5:00 pm**
- On **09 June 2011**: Be prepared to present your solution
 - English (or German)
 - Absent: Send us an e-mail in advance

Thanks for Listening

- Updates
 - Mailing list: searchengines2011 (at) hpi (...)
 - See website
- Questions
 - Via e-mail:
 - dustin (dot) lange (at) hpi (...)
 - saeedeh (dot) momtazi (at) hpi (...)
 - Office: A-1.6 / A-1.7

