

Salient terms extraction: Multimedia Image Retrieval for News Articles

Co-referencing entities, Boosting, Text-Rank

Jayaprakash Sundararaj,
Academic Researcher at Multimedia Search Team,
July, 2013

Agenda

- Goal (3:00 min)
- Challenges (3:00 min)
- Experiments (8:00 min)
- Future Direction (2:00 min)
- Acknowledgements (2:00 min)
- Beyond Work (2:00 min)

Goal

App of the Week: Muzio

App Name: Muzio Available Platforms: iOS, Desktop Price: Free What does this app do?: Reshma Chamberlin and Elizabeth Buchanan share a lot in common.



App of the Week: Muzio

App Name: Muzio Available Platforms: iOS, Desktop Price: Free What does this app do?: Reshma Chamberlin and Elizabeth Buchanan share a lot in common.

Overview

Input News Article &
Title

Choose candidate
keywords

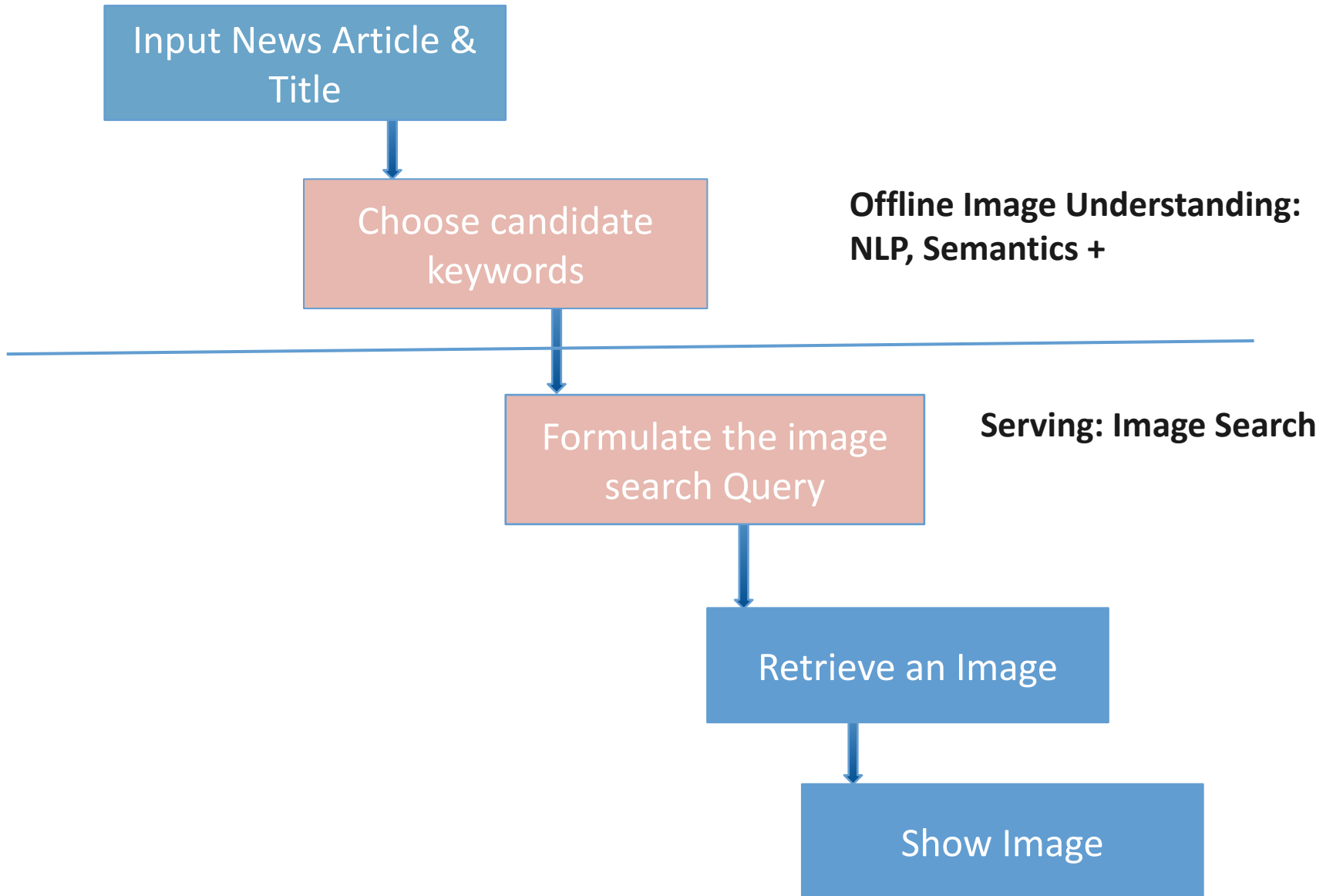
**Offline Image Understanding:
NLP, Semantics +**

Formulate the image
search Query

Serving: Image Search

Retrieve an Image

Show Image



Challenges

- Is **title** isn't enough?
 - example article title: '*Loosing Jobs*'
 - What does 'jobs' refers to?
 - Is it Steve Jobs?
 - Is it referring workplace firing?
- Is using **complete article** is too much of text?
 - Articles have 200 to 500 words
 - It is too expensive to use all of them for retrieval

Co-Reference

- One simple method of ranking noun phrases/entities could be based on how many times they are referred in the text.
- Entities are not referred in the same manner throughout the article. In most of the articles either part of names, pronouns, acronyms used to refer the same.
- *Example: Steve Jobs can be referred as Steve, Jobs, he, his, him... etc.*
- To identify the above in the article *Stanford Co-reference Pipeline* is used.

Co-Reference

- Boosting based on where noun-phrase appears in an article.
- If we consider noun phrase ‘custom duties on Chinese solar panels’ and later references to ‘duties’ indirectly refers to Chinese solar panels. So each reference adds boost to ‘Chinese solar panel’ too.

Sample article

The European Union says it will impose anti-dumping levies on imports of Chinese solar panels in a move that could trigger a trade war between two of the world's largest economies.

The 27-nation bloc will impose a tariff of about 12 percent on the import of solar panels, cells and wafers this week, EU Trade Commissioner Karel De Gucht announced Tuesday. That will increase to an average of 47 percent in August unless a settlement is reached with China in the next 60 days.

China the world's largest producer of solar panels, is accused by the EU of dumping them, or selling at improperly low prices, and hurting European competitors. Its exports of solar panels to Europe totaled 21 billion euros in 2011.

....

PageRank on Noun Phrases

- Purpose is to tryout scoring methods other than frequency of entities and it is based on how surrounding entities are connected.
- Two consecutive nouns phrases in a sentences are connected.
Example: In the sentence “EU will impose anti-dumping levies.”.
‘EU’ and ‘anti-dumping levies’ are connected.
- Two noun phrases/entities are connected, If they are co-referred.
Example in two sentences “Ramu is a Student.” and “He is intelligent..” here Ramu and He are connected.

Node Score

- Node's score in the Graph is calculated by:

$$W_Score(V_i) = (1 - d) + d * \sum_{V_j \in N_{gb}rs(V_i)} \frac{W_Score(V_j)}{\deg(V_j)}$$

- Idea is based on paper '*Rada Mihalcea, Paul Tarau, 'TextRank: Bringing Order into Texts'*'.

The European Union says it will impose anti-dumping levies on imports of Chinese solar panels, in a move that could trigger a trade war between two of the world's largest economies.

The 27-nation bloc will impose a tariff of about 12 percent on the import of solar panels, cells and wafers this week, EU Trade Commissioner Karel De Gucht announced Tuesday. That will increase to an average of 47 percent in August unless a settlement is reached with China in the next 60 days.

China, the world's largest producer of solar panels, is accused by the EU of dumping them, or selling at improperly low prices, and hurting European competitors. Its exports of solar panels to Europe totaled 21 billion euros in 2011.

Example

At least 7 injured in building explosion in Lower Manhattan

Yahoo! News Jul 11 11:26am

At least seven people have been **injured** — including 2 critically — after an explosion in New York City's Chinatown area.

FROM CAP ALGORITHM

1. Chinatown
 2. Explosion
 3. Lower Manhattan
 4. Christopher Kui
 5. Building
1. New York Fire Department
 2. Fire Department
 3. Fire Department spokesman
 4. Manhattan Bridge
 5. Christopher Kui
 6. New York Presbyterian Hospital

Comparison with CAP

- Performs better if important phrases repeated many times
- How is different from CAP ?
 - Detects 'Person, Location, Organization' well. They are ranked higher.
 - Events like 'The explosion', 'train accident' are poorly detected, which are important

Future Directions

- Semantic Role Labeling the parts of sentences with **PERSON**, **LOCATION**, **ORGANIZATION**, **DEVICE**, **TIME**, **EVENT**,, etc
- Giving important to the certain type of phrases, with respect to topic of Article:
 - ‘Event , Location’ for Casualties
 - ‘Apps, Device, Organization’ for Technology
 - ‘Named Entities, ’ for Celebrities

References

- Rada Mihalcea, Paul Tarau, TextRank: Bringing Order into Text
- Lawrence Page, Method for node ranking in a linked database
- Marta Recasens, Marie-Catherine de Marneffe, and Christopher Potts. 2013. The Life and Death of Discourse Entities: Identifying Singleton Mentions. In Proceedings of NAACL 2013.
- Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint Entity and Event Coreference Resolution across Documents. In Proceedings of the Conference on EMNLP-CoNLL.
- Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, Christopher Manning. 2010. A Multi-Pass Sieve for Coreference Resolution, EMNLP-2010

Thank you

Questions?

Acknowledgement

- Krishanu Seal, Senior Director
- Prateeksha Uday Chandraghatgi, Principal Engineer
- Kiran Kumar Palukuri, Senior Manager
- Abhranil Chatterjee, Senior Engineer
- Swapna M, Senior Engineer
- Bijoy Kumar Sarkar, Senior Engineer
- Hari A V, Senior Engineer
- Girish Krishna Ananthakrishnan, Staff Engineer
- Adarsh T, Hiring

APPENDIX

Beyond Work

- Fun with Team
- Yahoo Summer School - Information Retrieval and Semantic Web
- Got to know Awesome Projects