

WebTLab: A co-occurrence-based approach to KBP 2010 Entity-Linking task

Prof. Luis Sanchez-Fernandez
Web Technologies Laboratory
University Carlos III of Madrid
<http://webtlab.it.uc3m.es>



Outline

- ▣ Introduction
- ▣ Strategy
- ▣ Outline of the system
- ▣ Results
- ▣ Conclusions

Introduction

New features Log in / create account



Main page
Contents
Featured content
Current events
Random article
Donate

Interaction
2009-09-17
The most po
creative directi

Article Discussion Read Edit View history

Search

Search

Kim Basinger

From Wikipedia, the free encyclopedia
(Redirected from *Kim Bassinger*)

Kimila Ann "Kim" Basinger

(pronounced /ˈbeɪsɪŋər/ *BAY-sing-ər*, often mispronounced /ˈbæsɪndʒər/ *BASS-in-jər*; born December 8, 1953) is an American actress and former fashion model.

She is known for her portrayals of *Domino Petachi*, the Bond girl in *Never Say Never Again* (1983), and *Vicki Vale*,

Kim Basinger



საქართველო
Sakartvelo
Georgia^[1]



her hand at

ESSENCE.COM: How's your book tour going? What has the feedback been from women who have read it thus far?

NENE LEAKES: The turnout at all the signings has been amazing. I've been averaging about 300 people at most of them. In my hometown of **Athens, Georgia**, I received a very warm welcome—over 400 people came out for the signing. I'm a first-time author, but I've heard that's really good. Women have said they know me better from reading the book. I've gotten nothing but love.

ESSENCE.COM: This season on "The Real Housewives of **Atlanta**" you reunited with **Kim**. What do you have to say to those viewers who think you guys are fake because you weren't cool when the show ended last season?

LEAKES: Well, it's not fake. We're just like any other group of girls that hang out. We have our ups and downs, we get back together—that's real life. There's nothing on the show that is fake. It is what it is.

ESSENCE.COM: You directed an alter-ego photo shoot in an episode this season which we all loved. Do you have the bug to do more creative directing down the road?

LEAKES: I'm a reality star, I'm an author, I'm a businesswoman and I'm a director. I really like the entertainment industry. Directing? I love that kind of stuff. In tonight's episode, I will reveal the final photos from the shoot to all of the ladies. It is difficult to direct people because everyone doesn't see your vision, but at the end I think they all appreciated what I did for them. The job was definitely meant for me.

- ☐ FirstPerson
 - ☐ Identifier
 - ☐ JobTitle
 - ☒ Location
 - ☐ Lookup
 - ☒ Person
 - ☐ Sentence
 - ☐ SpaceToken
 - ☐ Split
 - ☐ Temp
 - ☐ Token
 - ☐ Unknown
- Original markups

Strategy I

WIKIPEDIA The Free Encyclopedia

Article Discussion

Atlanta

From Wikipedia, the free encyclopedia

This article (disambiguation) lists several topics that share the name Atlanta.

Atlanta (pronounced /ətˈlæntə/, /ɑːtlæntə/, colloquially /æ/) is a city in the state of Georgia, United States. It is the largest city in the Southeastern United States and the 35th largest city in the United States. Atlanta is the capital of Georgia and the seat of DeKalb County. It is also the seat of the Atlanta-Fulton County Metropolitan Area, which is the 12th largest metropolitan area in the United States. Atlanta is a major center for commerce, industry, and culture. It is home to several major universities, including Morehouse College, Spelman College, and Clark College. Atlanta is also a major hub for the airline industry, with Hartsfield-Jackson Atlanta International Airport being the busiest airport in the world.

NYC Gossip Girl
2009-09-17T23:38:00
The most popular housewife of Atlanta gets creative directing? Check out what Nene had to say to Essence below.

ESSENCE.COM: How's your book tour going? What has the feedback been from women who have read it thus far?

NENE LEAKES: The turnout at all the signings has been amazing. I've been averaging about 300 people at most of them. In my hometown of Athens, Georgia, I received a very warm welcome—over 400 people came out for the signing. I'm a first-time author, but I've heard that's really good. Women have said they know me better from reading the book. I've gotten nothing but love.

ESSENCE.COM: This season on "The Real Housewives of Atlanta" you reunited with Kim. What do you have to say to those viewers who think you guys are fake because you weren't cool when the show ended last season?

LEAKES: Well, it's not fake. We're just like any other group of girls that hang out. We have our ups and downs, we get back together—that's real life. There's nothing on the show that is fake. It is what it is.

ESSENCE.COM: You directed an alter-ego photo shoot in an episode this season which we all loved. Do you have the bug to do more creative directing down the road?

LEAKES: I'm a reality star, I'm an author, I'm a businesswoman and I'm a director. I really like the entertainment industry. Directing? I love that kind of stuff. In tonight's episode, I will reveal the final photos from the shoot to all of the ladies. It is difficult to direct people because everyone doesn't see your vision, but at the end I think they all appreciated what I did for them. The job was definitely meant for me.

WIKIPEDIA The Free Encyclopedia

Article Discussion

Athens

From Wikipedia, the free encyclopedia

Athens-Clarke County is a city-county in Georgia, northeastern part of the state. It is the home of the University of Georgia and is a college town. It is the initial growth of the city of Athens, which was founded in 1786. The city was the preceding year, the original city abandoned its charter in order to form a unified government with Clarke County.

Georgia (U.S. state)

From Wikipedia, the free encyclopedia

Coordinates: 33°N 83.5°W

For other uses, see Georgia.

Georgia (ⁱ/ˈdʒɔːrdʒə/) is a state located in the southeastern United States. Georgia was established in 1732, the last of the original Thirteen Colonies.

Georgia

☐ Date
☐ FirstPerson
☐ Identifier
☐ JobTitle
☒ Location
☐ Lookup
☒ Person
☐ Sentence
☐ SpaceToken
☐ Split
☐ Temp
☐ Token
☐ Unknown
► Original markups

Strategy II

▣ Approach

- ▣ Find entities in document
- ▣ For each entity, identify candidate instances that are compatible with the entity name
- ▣ Assign a ranking value to each candidate instance: $0 \leq r \leq 1$
 - ▣ Greater ranking values indicate greater likelihood of occurrence

Strategy III

- Semantic coherence (in terms of ranking)
 - “An instance would have a high ranking value if the instances that typically co-occur with it also have high ranking values”

$$r(I_i) = \sum_{j \in C_i} \text{Cooc}(I_i, I_j) r(I_j)$$

$$\vec{R} = A\vec{R}$$

<http://webtlab.it.uc3m.es>



Universidad
Carlos III de Madrid

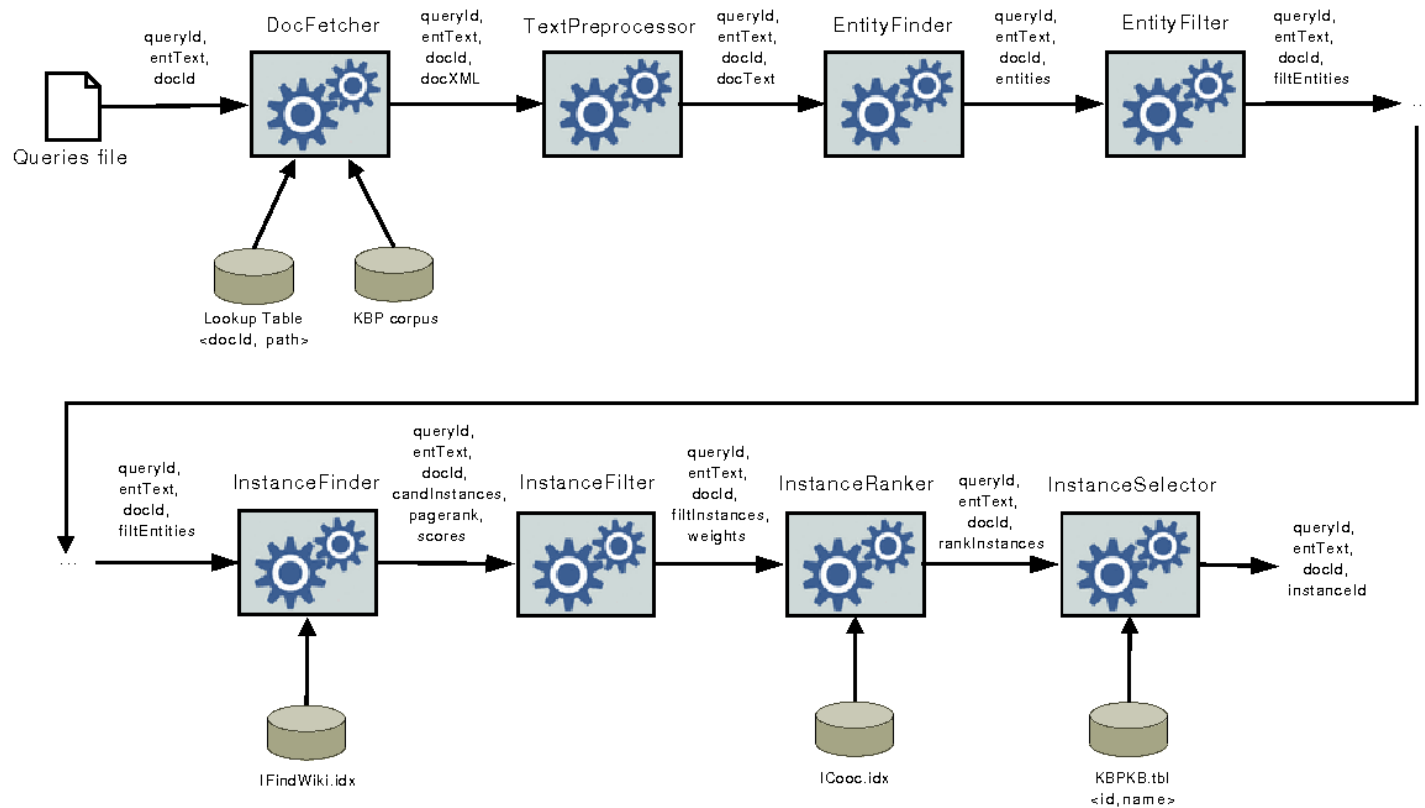


Strategy IV

- We can add a vector E that accounts for other context information
- Equation similar to Google PageRank

$$\vec{R} = (1 - \alpha) A \vec{R} + \alpha \vec{E}$$

Outline of the system



Entity finder & filter

- Use 3 open source tools
 - GATE (ANNIE): University of Sheffield
 - Named Entity Recognizer (NER): Stanford University
 - LbjNerTagger: University of Illinois at Urbana-Champaign
- Decide by majority
- Limit the number of entities to 100

Instance finder & filter

- Alternative instance names extracted by processing a Wikipedia dump
 - Page titles, redirects, disambiguation pages, anchors
 - Indexed by Lucene
- Candidate instances are obtained by querying Lucene
- Candidate instances weighted by combining Lucene scores and PageRank values
- Filtering limits the maximum number of candidates

Instance ranker

A_L : based on direct links
 A_C : based on instance weights passed by Wikipedia pager

The figure shows three Wikipedia pages illustrating the instance ranking process:

- Top Left:** Wikipedia page for "List of business schools in the United States". A blue arrow points from this page to the "Atlanta" article.
- Bottom Left:** Wikipedia page for "Atlanta". A blue arrow points from this page to the "Atlanta, Georgia" disambiguation page.
- Right:** Wikipedia page for "Athens, Georgia". A blue arrow points from this page to the "Athens-Clarke County, Georgia, USA" map.

$$\vec{R} = (k_L A_L \vec{R} + k_C A_C \vec{R}) + k_E \vec{E}$$

Instance selector

- If no candidate found: NIL
- If the first ranked candidate is not in the KB: NIL
- Otherwise, decide by comparing the two candidates with highest rank
 - Different threshold if exact match found

$$Plausibility = \frac{TopInstanceRanking}{SecondInstanceRanking}$$

Results I

Run	I. finder		I. ranker			I. selector	
	α_L	α_{PR}	k_L	k_C	k_E	σ_L	σ_H
WebTLab1	0.8	0.2	0.55	0.25	0.2	1.2	2.0
WebTLab2	0.8	0.2	0.55	0.25	0.2	1.05	1.5
WebTLab3	0.8	0.2	0.4	0.4	0.2	1.2	2.0

Run	2250 queries	1020 non-NIL	1230 NIL
WebTLab1	0.7698	0.6647	0.8569
WebTLab2	0.7636	0.6098	0.8911
WebTLab3	0.7596	0.6049	0.8878

$$\vec{R} = \left(k_L A_L \vec{R} + k_C A_C \vec{R} \right) + k_E \vec{E}$$

Results II

Run	I. finder		I. ranker			I. selector	
	α_L	α_{PR}	k_L	k_C	k_E	σ_L	σ_H
WebTLab1	0.8	0.2	0.55	0.25	0.2	1.2	2.0
WebTLab2	0.8	0.2	0.55	0.25	0.2	1.05	1.5
WebTLab3	0.8	0.2	0.4	0.4	0.2	1.2	2.0

Run	ORG	GPE	PER
WebTLab1	0.7613	0.6569	0.8908
WebTLab2	0.7707	0.6262	0.8935
WebTLab3	0.7680	0.6195	0.8908

$$\vec{R} = \left(k_L A_L \vec{R} + k_C A_C \vec{R} \right) + k_E \vec{E}$$

Results III: ablation tests

Run	I. finder		I. ranker			I. selector	
	α_L	α_{PR}	k_L	k_C	k_E	σ_L	σ_H
Lucene	1.0	0.0	0.0	0.0	1.0	1.2	2.0
Lucene-PR	0.8	0.2	0.0	0.0	1.0	1.2	2.0
WebTlab1	0.8	0.2	0.55	0.25	0.2	1.2	2.0

Run	2250 queries	1020 non-NIL	1230 NIL
Lucene	0.6364	0.2627	0.9463
Lucene-PR	0.6658	0.3363	0.9390
WebTlab1	0.7698	0.6647	0.8569

$$\vec{R} = \left(k_L A_L \vec{R} + k_C A_C \vec{R} \right) + k_E \vec{E}$$

Conclusions

- ▣ Approach based on instance co-occurrence
- ▣ Text from Wikipedia restricted to: titles, anchors
- ▣ Results considered promising
 - ▣ Should improve for GPE

Thank You!

Questions?