

Part I

Entity Linking

Outline

- Part 1 – Entity Linking
 - introduction
 - methods
 - evaluation
 - datasets
 - open challenges
 - hands-on
- Part 2 – Entity Retrieval

Introduction

article discussion edit this page history

You're running!

Plant

From Wikipedia, the free encyclopedia

For other uses, see *Plant* (disambiguation)

Plants are a major group of living things including familiar organisms such as trees, flowers, herbs, ferns, and mosses.

About 350,000 species of plants, defined as seed plants, bryophytes, ferns and fern allies, have been estimated to exist. As of 2004, some 287,655 species had been identified, of which 258,650 are flowering and 15,000 bryophytes.

Tree

From Wikipedia, the free encyclopedia

A tree is a large, perennial, woody plant. Though there is no set definition regarding minimum size, the term generally applies to plants at least 6 m (20 ft) high at maturity and, more importantly, having



Fossil range: Middle-Late Ordovician - Recent



Species

From Wikipedia, the free encyclopedia

This article is about biology. For the movie, see Species (film). In biology, a **species** is one of the basic units of biodiversity. In classification, a species is assigned a two-part name, treated as a binomial nomenclature. The first part is the genus, followed by the specific epithet. For example, humans belong to the genus *Homo*, and the species *Homo sapiens*. The name of the species is the whole name, just the second term (which may be called *specific epithet*).

Image taken from Mihalcea and Csomai (2007). **Wikify!: linking documents to encyclopedic knowledge.** In CIKM '07.

**Let's learn something about
Spin-Optical Metamaterial**

Spin-Optical Metamaterial Route to Spin-Controlled Photonics

www.sciencemag.org/content/340/6133/724

Spin-Optical Metamaterial Route to Spin-Controlled Photonics

Science AAAS NEWS SCIENCE JOURNALS CAREERS BLOGS & COMMUNITIES MULTIMEDIA COLLECTIONS JOIN / SUBSCRIBE

Science The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

Science Home Current Issue Previous Issues Science Express Science Products My Science About the Journal

Home > Science Magazine > 10 May 2013 > Shitrit et al., 340 (6133): 724–726

Article Views Abstract Full Text Full Text (PDF) Figures Only Supplementary Materials

Science 10 May 2013: Vol. 340 no. 6133 pp. 724–726 DOI: 10.1126/science.1234892

REPORT

Spin-Optical Metamaterial Route to Spin-Controlled Photonics

Nir Shitrit, Igor Yulevich, Elhanan Maguid, Dror Ozeri, Dekel Veksler, Vladimir Kleiner, Erez Hasman*

*Corresponding author. E-mail: mehasman@technion.ac.il

ABSTRACT EDITOR'S SUMMARY

Spin optics provides a route to control light, whereby the photon helicity (spin angular momentum) degeneracy is removed due to a geometric gradient onto a metasurface. The alliance of spin optics and metamaterials offers the dispersion engineering of a structured matter in a polarization helicity-dependent manner. We show that polarization-controlled optical modes of metamaterials arise where the spatial inversion symmetry is violated. The emerged spin-split dispersion of spontaneous emission originates from the spin-orbit interaction of light, generating a selection rule based on symmetry restrictions in a spin-optical metamaterial. The inversion asymmetric metasurface is obtained via anisotropic optical antenna patterns. This type of metamaterial provides a route for spin-controlled nanophotonic applications based on the design of the metasurface symmetry properties.

Received for publication 7 January 2013.
Accepted for publication 13 March 2013.

ADVERTISEMENT

Science MOBILE Now Available for Android Phones

Scan the barcode to download from the Android Market.

ADVERTISEMENT

WOMEN IN SCIENCE forging new pathways in biology
16 inspiring profiles

Display a menu for "http://oascentral.sciencemag.org/RealMedia/ads/click_lx.ads/www.sciencemag.org/content/340/6133/724/L18/717884666/Right1/AAAS/SOL-CAR-HouseAdsAndroidApp101201/J-2957_MobileApp-Android_160x150b_v1.html/306c72594d56474c34783441427"

REPORT



Spin–Optical Metamaterial Route to Spin–Controlled Photonics

Nir Shitrit, Igor Yulevich, Elhanan Maguid, Dror Ozeri, Dekel Veksler, Vladimir Kleiner, Erez Hasman*

Author Affiliations

*Corresponding author. E-mail: mehasman@technion.ac.il

ADV

ABSTRACT

EDITOR'S SUMMARY

Spin optics provides a route to control light, whereby the photon helicity (spin angular momentum) degeneracy is removed due to a geometric gradient onto a metasurface. The alliance of spin optics and metamaterials offers the dispersion engineering of a structured matter in a polarization helicity-dependent manner. We show that polarization-controlled optical modes of metamaterials arise where the spatial inversion symmetry is violated. The emerged spin-split dispersion of spontaneous emission originates from the spin-orbit interaction of light, generating a selection rule based on symmetry restrictions in a spin-optical metamaterial. The inversion asymmetric metasurface is obtained via anisotropic optical antenna patterns. This type of metamaterial provides a route for spin-controlled nanophotonic applications based on the design of the metasurface symmetry properties.



WC
IN SC
forgi
patl
in h

Received for publication 7 January 2013.

Input Text

 Italiano  English

momentum) degeneracy is removed due to a geometric gradient onto a metasurface. The alliance of spin optics and metamaterials offers the dispersion engineering of a structured matter in a polarization helicity-dependent manner. We show that polarization-controlled optical modes of [metamaterials](#) arise where the spatial inversion symmetry is violated. The emerged spin-split dispersion of spontaneous emission originates from the spin-orbit interaction of light, generating a selection rule based on symmetry restrictions in a spin-optical metamaterial. The inversion asymmetric metasurface is obtained via anisotropic optical antenna patterns. This type of metamaterial provides a route for spin-controlled nanophotonic applications based on the design of the metasurface symmetry properties.

Many links

Few links

Reset

TAGME!

Tagged text Topics

Spin [optics](#) provides a route to [control light](#), whereby the [photon helicity](#) (spin [angular momentum](#)) [degeneracy](#) is removed due to a [geometric gradient](#) onto a metasurface. The alliance of sp **Degenerate energy levels**
In physics, two or more different quantum states are said to be degenerate if they are all at the same energy level. Statistically this means that they are all equally probable of being filled, and in...
e [dispersion engineering](#) of a structured matter. We show that polarization-controlled [spatial inversion symmetry](#) is violated. The [emission](#) originates from the [spin-orbit](#) interaction based on [symmetry](#) restrictions in a spin-metasurface is obtained via [anisotropic](#) optical antenna patterns. This [type](#) of metamaterial provides a route for spin-controlled

See <http://tagme.di.unipi.it>



Degenerate energy levels – Wikipedia, the free encyclopedia

W en.wikipedia.org/wiki/Degenerate_energy_level

Degenerate energy levels – Wikipedia, the free encyclopedia

Input Text

Italiano English

Degenerate energy levels

From Wikipedia, the free encyclopedia
(Redirected from [Degenerate energy level](#))

This article is about different quantum states having the same energy. For other uses, see [Degeneracy](#).
"Quantum degeneracy" redirects here. It sometimes refers to a [degenerate matter](#).

This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (February 2009)

In quantum mechanics, a branch of physics, two or more different states of a system are said to be degenerate if they are all at the same energy level. It is represented mathematically by the system having more than one linearly independent eigenstate with the same eigenvalue. Conversely, an energy level is said to be degenerate if it contains two or more different states at a particular energy level is called the level's degeneracy, and this phenomenon is generally known as a quantum degeneracy.

From the perspective of quantum statistical mechanics, several degenerate states at the same level are all equally probable of being filled.

Contents [hide]

- 1 Mathematics
- 2 Examples
- 3 Perturbation
- 4 See also
- 5 Further reading

Mathematics

The term comes from the fact that, for a point spectrum Hamiltonian H , degenerate eigenstates correspond to identical eigenvalues. Since eigenvalues correspond to roots of the characteristic polynomial, the word degeneracy here has the same meaning as the common mathematical usage of the word.

The eigenvalue λ is called nondegenerate (or simple) when its corresponding eigenvector is unique up to a constant factor, or, the same, the corresponding eigenspace is one-dimensional. Indeed, the eigenspace $\{\psi : H|\psi\rangle = \lambda|\psi\rangle\}$ (in bra-ket notation) is not necessarily one-dimensional. If there exist at least two linearly independent ket-vectors in it, then this eigenvalue is called degenerate. Its degree of degeneracy is then the dimension of the eigenspace, which is the same as the number of distinct (linearly independent) quantum states associated with it.

Examples

In atomic physics, electron's energy levels are often degenerate, where different possible occupation states for particles may be related by symmetry. For example, in the hydrogen atom, for a given principal quantum number n , there exist several states which have that energy, but differ in the eigenvalues of angular momentum L^2 , spin component S_z and so on. The eigenvalue of an operator which is zero for all degenerate states is called a quantum number.

See <http://tagme.di.unipi.it>



Office

HOME MY OFFICE PRODUCTS SUPPORT IMAGES TEMPLATES STORE

Search all of Office.com

About smart tags

You can save time by using smart tags to open other programs to do.

The purple dotted lines beneath text indicate where a smart tag has been applied.

Nate Sun dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut Steve Knopf magna aliquam erat volutpat.

① Smart tag indicators

- + How to use smart tags
- + How smart tags work
- + How to get more smart tags
- + Smart tag options
- + Creating smart tags and setting them up

Smart tagged text in Word

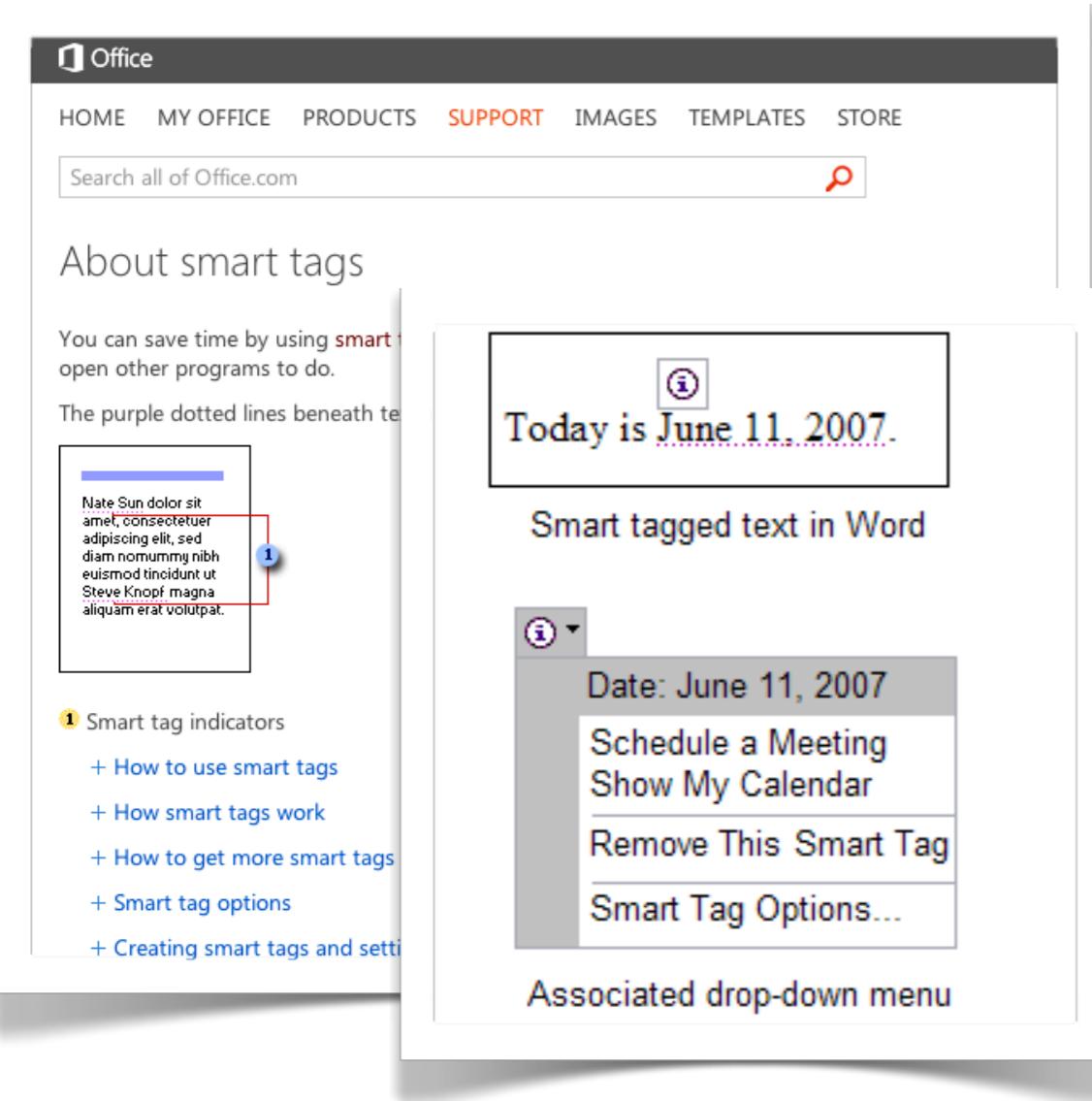
Date: June 11, 2007

Schedule a Meeting
Show My Calendar

Remove This Smart Tag

Smart Tag Options...

Associated drop-down menu



Microsoft Smart Tags

http://cse.unl.edu/~choueiry/S01-476-876/

Google

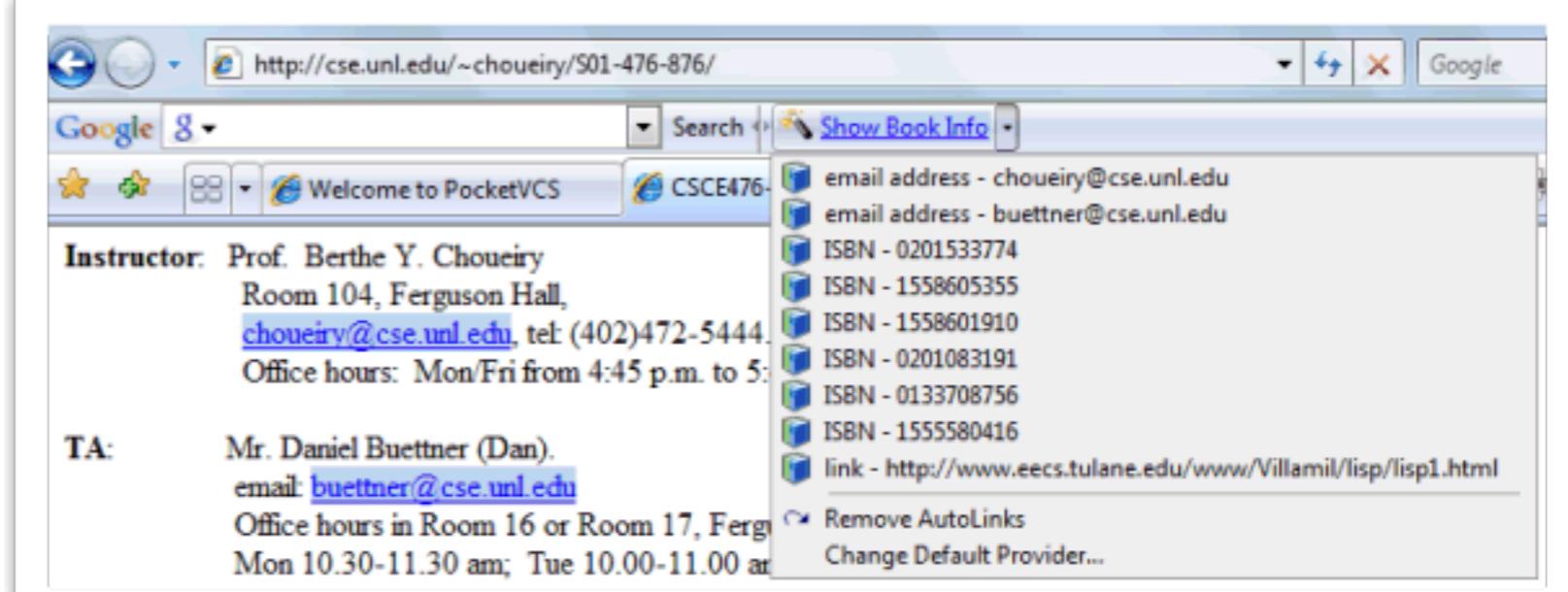
Welcome to PocketVCS CSCE476

Instructor: Prof. Berthe Y. Choueiry
Room 104, Ferguson Hall,
choueiry@cse.unl.edu, tel: (402)472-5444.
Office hours: Mon/Fri from 4:45 p.m. to 5:30 p.m.

TA: Mr. Daniel Buettner (Dan).
email: bueettner@cse.unl.edu
Office hours in Room 16 or Room 17, Ferguson Hall
Mon 10.30-11.30 am; Tue 10.00-11.00 am

email address - choueiry@cse.unl.edu
email address - buettner@cse.unl.edu
ISBN - 0201533774
ISBN - 1558605355
ISBN - 1558601910
ISBN - 0201083191
ISBN - 0133708756
ISBN - 1555580416
link - http://www.eecs.tulane.edu/www/Villamil/lisp/lisp1.html

Remove AutoLinks
Change Default Provider...



Google toolbar

Bing

www.bing.com/search?q=rio+de+janeiro&go=&qs=n&form=QBLH&filt=all&pq=rio+de+janeiro&sc=8-14&sp=-1&sk=

Reader

rio de janeiro – Bing

WEB IMAGES VIDEOS SHOPPING NEWS LOCAL MORE

bing Beta Edgar

28.500.000 RESULTS Narrow by language ▾ Narrow by region ▾

250 Hoteles Río Janeiro
booking.com/Rio-De-Janeiro-Hoteles
Reserva online: rápido y seguro. Reservar un Hotel en Río de Janeiro.

Río De Janeiro desde 34€
Rio-De-Janeiro.eDreams.es
Dto. en Vuelos a Río De Janeiro: Hasta 70% de descuento, ¡Reserva Ya!

Images of rio de janeiro
bing.com/images

Río de Janeiro - Wikipedia, la enciclopedia... Translate this page
es.wikipedia.org/wiki/Río_de_Janeiro ▾
Toponimia · Historia · Demografía · Geografía · Política · Economía
Río de Janeiro (pronunciado AFI: [ˈrl̥.u dʒi ʒaˈnejɾu] en el portugués brasileño culto, fundada como São Sebastião do Río de Janeiro , «San Sebastián del ...

Guía de Río de Janeiro - guiarte.com... Translate this page
www.guiarte.com/rio-de-janeiro ▾
Guía de Río de Janeiro, una de las ciudades más hermosas de América: Río de Janeiro.

Videos of rio de janeiro
bing.com/videos

Río de Janeiro Brasil Michael Mind - Rio De Janeiro 80 Discover Rio de Janeiro - Brasil

Display a menu

Bing

www.google.com/#output=search&sclient=psy-ab&q=rio+de+janeiro&oq=rio+de+janeiro&gs_l=hp.3...15914.15914.0.16069.1.1.0.0.0.0.0.0.0...0.0...0.0...1c.2.12.psy-ab.rBzL4X8pUk&pbx= Reader

Edgar Meij 0 + Share

Google rio de janeiro

Web Images Maps Shopping News More Search tools

About 362,000,000 results (0.27 seconds)

Rio de Janeiro - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Rio_de_Janeiro ▾
Rio de Janeiro (pron.: /rɪou̯ də̯ʒənɛ̯ɛ̯r̯o̯/ or /rɪou̯ dʒə̯nɛ̯ɛ̯r̯o̯/; Portuguese pronunciation: [ˈri.u̯ dʒi ʒɐ̯nɛ̯ru̯], January River), commonly ...
Christ the Redeemer - Rio de Janeiro - History of Rio de Janeiro - Avenida Brasil

Rio de Janeiro travel guide - Wikitravel
wikitravel.org/en/Rio_de_Janeiro ▾
Open source travel guide to Rio de Janeiro, featuring up-to-date information on attractions, hotels, restaurants, nightlife, travel tips and more. Free and reliable ...

Images for rio de janeiro - Report images



Rio de Janeiro Tourism and Vacations: 522 Things to Do in Rio de ...
[www.tripadvisor.com/.../Brazil/State-of-Rio-de-Janeiro-\(RJ\).html](http://www.tripadvisor.com/.../Brazil/State-of-Rio-de-Janeiro-(RJ).html) ▾
★★★★★ Review by TripAdvisor Travelers' Choice
With breathtaking views from Corcovado Mountain and breathtaking deals in the city's endless malls and markets, Rio de Janeiro is a vacation paradise, ...

WWW 2013 – Rio de Janeiro, Brazil
www2013.org/ ▾
Rio, 13th-17th May. The World Wide Web Conference is a yearly international conference on the topic of the future direction of the World Wide Web.
You've visited this page 3 times. Last visit: 3/25/13

News for rio de janeiro



Display a menu


©2013 Google Map data ©2013 Google, MapLink

Rio de Janeiro

City in Brazil

Rio de Janeiro, commonly referred to simply as Rio, is the capital city of the State of Rio de Janeiro, the second largest city of Brazil, and the third largest metropolitan area and agglomeration in ... [Wikipedia](#)

Population: 6.323 million (2010)
Area: 486 sq miles (1,260 km²)
Weather: 25°C, Wind S at 19 km/h, 61% Humidity
Local time: Tuesday 2:41 PM

Points of interest


Copacabana Ipanema Cristo Redentor Leblon Sugarloaf Mountain

Feedback / More info

Google

rio de janeiro

New User? Register | Sign In

Make Y! My Homepage

Bucket: VIP048 Local BE Test #... bad results / ads or bugs? tell us! [hide]

YAHOO!

rio de janeiro

Search

Options

RELATED SEARCHES

- rio de janeiro olympics
- rio de janeiro carnival
- rio de janeiro beaches
- brazil
- sao paulo

FILTER BY TIME

- Anytime
- Past day
- Past week
- Past month

rio de janeiro

RIO DE JANEIRO, BRAZIL
02:41 PM (America/Sao_Paulo). - Current local time

TOP RATED THINGS TO DO

1. Corcovado
2. Pão de Açúcar
3. Arcos da Lapa

RIO DE JANEIRO OVERVIEW

Hotels

Restaurant Guide

Flights

travel.yahoo.com
Tue May 07 2:41 pm (BST) | Partly Cloudy, 77°F ☁
Sitting on the southern shore of the magnificent Guanabara Bay, RIO DE JANEIRO has, without a shadow of a doubt, one of the most stunning settings in the world. Extending for 20km along an alluvial strip, between an azure sea and forest-clad mountains, the city's streets and buildings have been moulded ...
[more](#)




Sugar Loaf Moun...

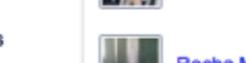

Resende, Brazil


Carnaval


Santana


Saúde


HSBC Arena


Rocha Miranda

Ads

VIP048
YI Confidential

[hide]

Display a menu

Bucket: VIP048 Local BE Test #... bad results / ads or bugs? tell us! [hide]

Yahoo!

rio de janeiro

New User? Register | Sign In

Make Y! My Homepage

Bucket: VIP048 | Local BE Test #... bad results / ads or bugs? tell us! [hide]

YAHOO!

rio de janeiro

Search Options

122,000,000 results

WEB IMAGES VIDEO SHOPPING BLOGS MORE

RELATED SEARCHES

- rio de janeiro olympics
- rio de janeiro carnival
- rio de janeiro beaches
- brazil
- sao paulo

FILTER BY TIME

- Anytime
- Past day
- Past week
- Past month

Also try: [rio de janeiro olympics](#), [rio de janeiro carnival](#), [rio de janeiro beaches](#), [more...](#)

Ad related to rio de janeiro

Flights to Rio de Janeiro

Fly TAM to Rio de Janeiro and Save. Great Fares and No Booking Fees!
www.tam.com.br/USA

More sponsors: [rio de janeiro](#), [flights](#)

Rio De Janeiro, Brazil

travel.yahoo.com
Tue May 07 2:41 pm (BST) | Partly Cloudy, 77°F ☁
Sitting on the southern shore of the magnificent Guanabara Bay, RIO DE JANEIRO has, without a shadow of a doubt, one of the most stunning settings in the world. Extending for 20km along an alluvial strip, between an azure sea and forest-clad mountains, the city's streets and buildings have been moulded ...
[more](#)



Rio de Janeiro - Wikipedia, the free encyclopedia

Geography | Climate | Cityscape | History
Rio de Janeiro, commonly referred to simply as Rio, is the capital city of the State of Rio de Janeiro, the second largest city of Brazil, and the third largest ...
en.wikipedia.org/wiki/Rio_de_Janeiro - Cached
[More results from en.wikipedia.org »](#)

Rio de Janeiro travel guide - Wikitravel

Rio de Janeiro is the second largest city in Brazil, on the South Atlantic coast. Rio is famous for its breathtaking landscape, its laidback beach culture and its ...
wikitravel.org/en/Rio - Cached
[More results from wikitravel.org »](#)

Rio De Janeiro - Image Results

Ads VIP048 Y! Confidential [hide]

Display a menu

Yahoo!

The screenshot shows the Yahoo! homepage with a purple header bar. The top navigation bar includes links for Mail, News, Finance, Sports, Movies, omg!, Shine, Autos, Shopping, Travel, Dating, and Jobs. Below the header is a large banner featuring two cheerleaders with the headline "Writer under fire for slamming cheerleader's weight". A sub-headline states: "A blogger says an Oklahoma City dancer has no business wearing a tiny outfit in front of an NBA crowd. She politely fires back »". Below the banner is a horizontal news strip with five thumbnails: "Blogger calls out cheerleader", "Paltrow's dress defended", "Paris Jackson with her mom", "Progressive Insurance lady", and "Michael Jordan marries". The main content area features several news stories: "Court may limit use of race in college admission decisions", "In a first, black voter turnout rate passes whites", "Dad Anticipates Tough Talks With His Teenage Daughters", and "Analysis: Israeli credibility on line over Iran nuclear challenge". On the right side, there is a "Trending Now" sidebar with ten items, a "YAHOO! AUTOS" section with a "Take a look" button and a car image, a weather forecast for London, and a footer with a red "Test H500 Y! Confidential" button.

Yahoo! Homerun

The screenshot shows the Yahoo! homepage with a purple header bar. The top navigation bar includes links for Mail, News, Finance, Sports, Movies, omg!, Shine, Autos, Shopping, Travel, Dating, and Jobs. Below the header is a large banner image of two cheerleaders. A headline reads "Writer under fire for slamming cheerleader's weight". A sub-headline says "A blogger says an Oklahoma City dancer has no business wearing a tiny outfit in front of an NBA crowd. She politely fires back »". Below the banner is a horizontal scrollable news strip with five thumbnails: "Blogger calls out cheerleader", "Paltrow's dress defended", "Paris Jackson with her mom", "Progressive Insurance lady", and "Michael Jordan marries". The main content area features several news stories: "In a first, black voter turnout rate passes whites" (Associated Press), "Dad Anticipates Tough Talks With His Teenage Daughters" (Dear Abby), and "Analysis: Israeli credibility on line over Iran nuclear challenge". On the right side, there are sections for "Trending Now" (with links to Eastwood age 105, Swift \$17 million man..., etc.), "YAHOO! AUTOS" (with a "Take a look" button and a row of cars), and a weather forecast for London (52°F Fair). The bottom of the page includes a "Display a menu" link and a red footer bar with icons for a warning sign, "Test H500 Y! Confidential", a speaker, and a bug.

Yahoo! Homerun

Goals of part I

- Learn entity linking basics
- Get familiar with
 - terminology and essentials
 - seminal papers/methods
 - evaluation and datasets
- Obtain hands-on experience with
 - (publicly available) toolkits
 - evaluation

Why?

- (Automatic) document enrichment
 - go-read-here
 - assistance for (Wikipedia) editors
 - inline (microformats, RDFa)
- Enable
 - semantic search
 - advanced UI/UX
 - ontology learning, KB population
- Improve
 - classification
 - word sense disambiguation, semantic similarity

A bit of history

- Text classification
- NER
- WSD
- NED/NEN
 - {person name, geo, movie name, ...} disambiguation
 - (Cross-document) coreference resolution
 - Automatic link generation
- Entity linking

Entity linking?

- NE normalization / canonicalization / sense disambiguation
- DB record linkage / schema mapping
- Knowledge base population
- Entity linking
 - D2W
 - Wikification
 - Semantic linking

Entity Linking: main problem

- Linking free text to *entities*
 - “things with a distinct and independent existence”
 - e.g., people, companies, products, locations, etc.
 - typically taken from a knowledge base
 - Wikipedia
 - Freebase
 - Text?
 - Anything
 - documents
 - tweets
 - queries
 - ...

Typical steps

1. Determine “linkable” phrases
 - mention detection – **MD**
2. Rank>Select candidate entity links
 - link generation – **LG**
 - may include NILs (null values, i.e., no target in KB)
3. (Use “context” to disambiguate/filter/improve)
 - disambiguation – **DA**

Methods

Preliminaries

- Wikipedia
- Wikipedia-based measures
 - commonness
 - relatedness
 - keyphraseness

Wikipedia

- Basic element: article (proper)
- But also
 - redirect pages
 - disambiguation pages
 - category/template pages
 - admin pages
- Hyperlinks
 - use “unique identifiers” (URLs)
 - [[United States]] or [[United States|American]]
 - [[United States (TV series)]] or
[[United States (TV series)|TV show]]



Some statistics

add stats
copy from <http://zookstl3.science.uva.nl:8080/dutchsemcor/getStats?>

- WordNet
 - 80k entity definitions
 - 115k surface forms
 - 142k senses (entity - surface form combinations)
- Wikipedia (only)
 - ~4M entity definitions
 - ~12M surface forms
 - ~24M senses

Wikipedia-based measures

Wikipedia-based measures

- keyphraseness(w) **[Mihalcea & Csomai 2007]**

$$\frac{\text{CF}(w_l)}{\text{CF}(w)}$$

Wikipedia-based measures

- keyphraseness(w) [Mihalcea & Csomai 2007]

$$\frac{\text{CF}(w_l)}{\text{CF}(w)} \longrightarrow \begin{array}{l} \textbf{Collection frequency} \\ \text{term } w \text{ as a link to another} \\ \text{Wikipedia article} \end{array}$$



Collection frequency
term w

Wikipedia-based measures

- commonness(w, c) [Medelyan et al. 2008]

$$\frac{|L_{w,c}|}{\sum_{c'} |L_{w,c'}|}$$



Number of links
with target c' and anchor text w

Commonness and keyphraseness

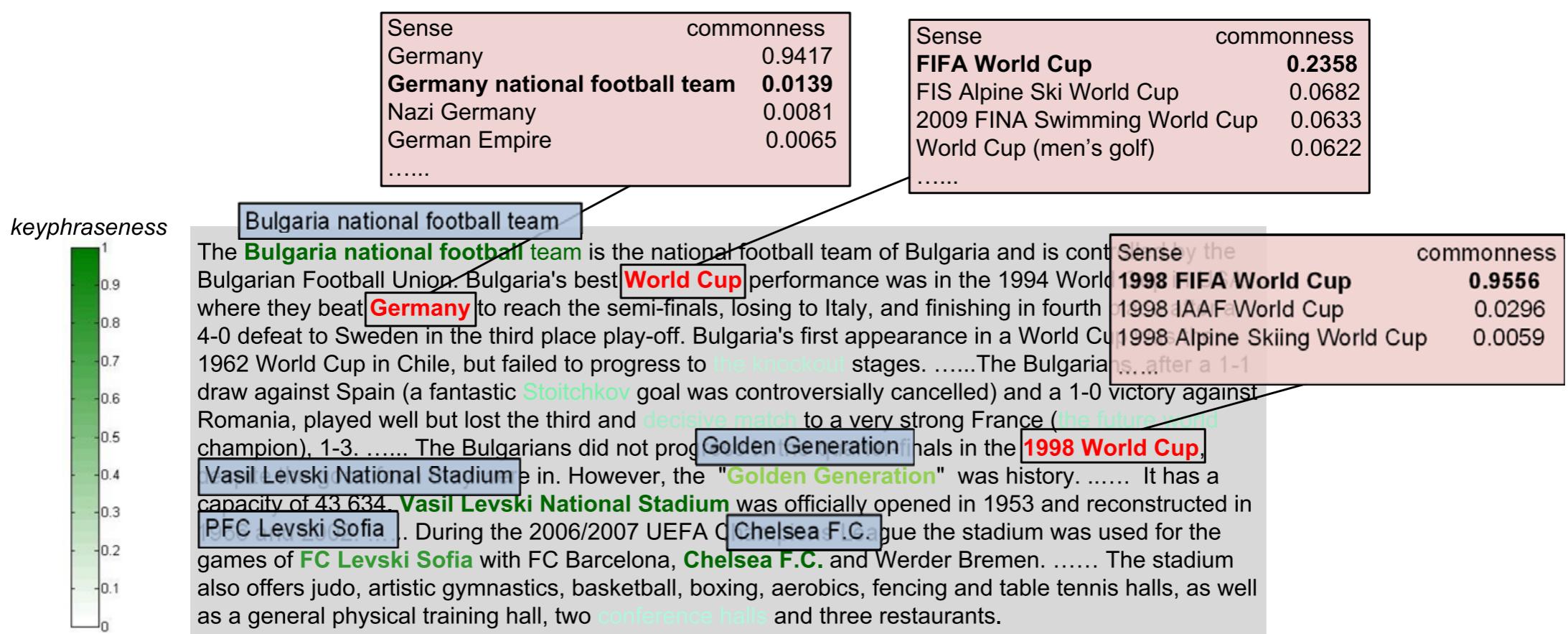


Image taken from Li et al. (2013). TSDW: Two-stage word sense disambiguation using Wikipedia. In JASIST 2013.

Wikipedia-based measures

- relatedness(c, c') [Milne & Witten 2008a]

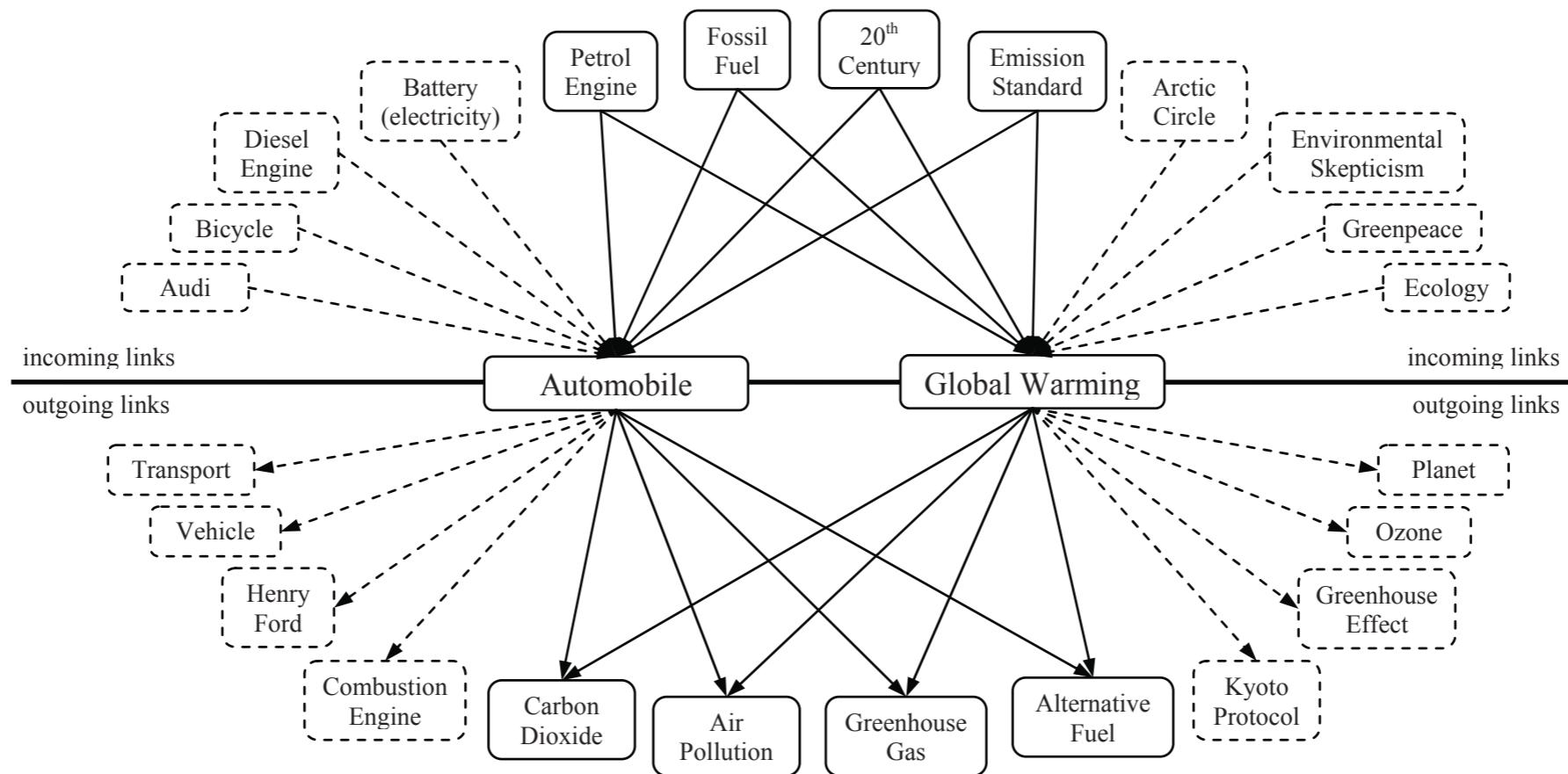


Image taken from Milne and Witten (2008a). **An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links.** In AAAI WikiAI Workshop.

Wikipedia-based measures

- relatedness(c, c') [Milne & Witten 2008a]

$$\frac{\log(\max(|L_c|, |L_{c'}|)) - \log(|L_c \cap L_{c'}|)}{\log(|WP|) - \log(\min(|L_c|, |L_{c'}|))}$$

Number of links with target c

Intersection of inlinks with target c and c'

Total number of Wikipedia articles

The diagram illustrates the formula for relatedness. It shows two upward arrows pointing to the terms $\log(\max(|L_c|, |L_{c'}|))$ and $\log(|L_c \cap L_{c'}|)$. The first arrow is associated with the label "Number of links with target c ". The second arrow is associated with the label "Intersection of inlinks with target c and c' ". A downward arrow points to the term $\log(|WP|) - \log(\min(|L_c|, |L_{c'}|))$, which is associated with the label "Total number of Wikipedia articles". Brackets group the terms $\log(\max(|L_c|, |L_{c'}|))$ and $\log(|L_c \cap L_{c'}|)$ under the denominator.

Baseline methods

Recall the steps

- 1. mention detection – MD**
- 2. link generation – LG**
- 3. (disambiguation) – DA**

Large-Scale Named Entity Disambiguation Based on Wikipedia Data

[Cucerzan 2007]

- Key intuition: leverage context links
 - ""**Texas**"" is a [[pop music]] band from [[Glasgow]], [[Scotland]], [[United Kingdom]]. They were founded by [[Johnny McElhone]] in [[1986 in music|1986]] and had their performing debut in [[March]] [[1988]] at ...
- Prune the candidates, keep only:
 - appearances in the first paragraph of an article, and
 - reciprocal links

Large-Scale Named Entity Disambiguation Based on Wikipedia Data

[Cucerzan 2007]

- MD
 - NER; rule-based; co-ref resolution
- LG
 - Represent entities as vectors
 - context, categories
 - Same for all candidate entity links
 - Determine maximally coherent set

Wikify!

[Mihalcea & Csomai 2007]

- MD
 - tf.idf, χ^2 , keyphraseness
- LG
 1. Overlap between definition (Wikipedia page) and context (paragraph) [Lesk 1986]
 2. Naive Bayes [Mihalcea 2007]
 - context, POS, entity-specific terms
 3. Voting between (1) and (2)

Topic Indexing with Wikipedia

[Medelyan et al. 2008]

- MD
 - keyphraseness [Mihalcea & Csomai 2007]
- LG
 - combination of average relatedness & commonness
- LG/DA
 - Naive Bayes
 - TF.IDF, position, length, degree, weighted keyphraseness

Learning to Link with Wikipedia

[Milne & Witten 2008b]

- Key idea: disambiguation informs detection
 - compare each possible sense with its *relatedness* to the context sense candidates
 - start with unambiguous senses

Learning to Link with Wikipedia

[Milne & Witten 2008b]

Depth-first search

From Wikipedia, the free encyclopedia



Depth-first search (DFS) is an algorithm for traversing or searching a tree structure or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	2.57%	63.26%
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

Image taken from Milne and Witten (2008b). Learning to Link with Wikipedia. In CIKM '08.

Learning to Link with Wikipedia

[Milne & Witten 2008b]

- Filter non-informative, non-ambiguous candidates (e.g., “the”)
 - based on keyphraseness, i.e., link probability
- Filter non-central candidates
 - based on average relatedness to all other context senses
- Combine

Learning to Link with Wikipedia

[Milne & Witten 2008b]

- MD
 - ...
- LG
 - Machine learning
 - keyphraseness, average relatedness, sum of average weights

Learning to Link with Wikipedia

[Milne & Witten 2008b]

- MD
 - Machine learning
 - link probability, relatedness, **confidence of LG**, generality, frequency, location, spread
- LG
 - Machine learning
 - keyphraseness, average relatedness, sum of average weights

Local and Global Algorithms for Disambiguation to Wikipedia

[Ratinov et al. 2011]

- Explicit focus on *global* versus *local* algorithms
 - “Global,” i.e., disambiguation of the candidate graph
 - NP-hard
- Optimization
 - reduce the search space to a “disambiguation context,” e.g.,
 - all plausible disambiguations [Cucerzan 2007]
 - unambiguous surface forms [Milne & Witten 2008b]

Local and Global Algorithms for Disambiguation to Wikipedia

[Ratinov et al. 2011]

- Main contribution, in steps
 1. use “local” approach (e.g., commonness) to generate a disambiguation context
 2. apply “global” machine learning approach
 - relatedness, PMI
 - {inlinks, outlinks} in various combinations (c and c')
 - {avg, max}
- Finally, apply another round of machine learning

TAGME: On-the-fly Annotation of Short Text Fragments

[Ferragina & Scaiella 2010]

- MD
 - keyphraseness [Mihalcea & Csomai 2007]
- LG
 - use “local” approach to generate a disambiguation context, similar to [Ratinov et al. 2011]
 - Heavy pruning
 - mentions; candidate links; coherence
- Accessible at <http://tagme.di.unipi.it>

Adding semantics to microblog posts

[Meij et al. 2012]

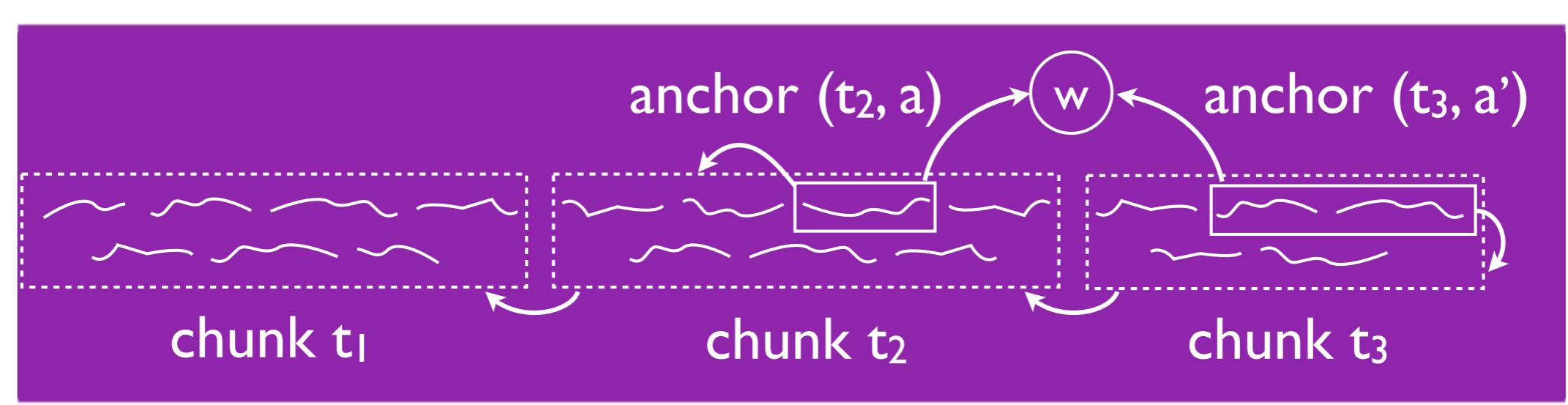
- MD
 - commonness (and others)
 - idea: keep ranked list of all candidate entity links
- LG
 - use machine learning to determine which of the links to keep
 - ..., random forests, GBRT
 - big set of {text, entity, text+entity, context} features

Graph-based methods

Feeding the Second Screen: Semantic Linking based on Subtitles

[Odijk et al. 2013]

- Setting: entity linking on closed captions
 - streaming, high-precision, real-time
- Graph information as additional features
 - Idea: maintain a (coherent) tripartite context graph
 - entities
 - chunks
 - anchors



Feeding the Second Screen: Semantic Linking based on Subtitles

[Odijk et al. 2013]

Context features

$DEGREE(w, G)$	Number of edges connected to the node representing Wikipedia article w in context graph G .
$DEGREE - CENTRALITY(w, G)$	Centrality of Wikipedia article w in context graph G , computed as the ratio of edges connected to the node representing w in G .
$PAGERANK(w, G)$	Importance of the node representing w in context graph G , measured using PageRank.

A Graph-based Method for Entity Linking

[Guo et al. 2011]

- MD
 - rule-based; prefer longer links
 - generate a disambiguation context
- LG
 - (weighted interpolation of) in- and outdegree in disambiguation context to select entity links
 - edges defined by wikilinks
- Evaluation on TAC KBP

Graph-based named entity linking with Wikipedia

[Hachey et al. 2011]

- MD
 - generate disambiguation context
 - based on unambiguous entity links
 - edges defined by wikilinks (articles & categories)
 - max step size: 2 (articles), 3 (categories)
- LG
 - use degree centrality and PageRank to reweigh cosine-based similarity scores
- Evaluation on TAC KBP

Recap

- Essential ingredients
 - MD
 - commonness
 - keyphraseness
 - LG
 - commonness
 - machine learning
 - DA
 - relatedness
 - machine learning

Evaluation

DIY Entity Linking

- Target KB (Wikipedia)
- Test collection
- Evaluation metrics

Measures

- Set-based (similar to WSD)
 - “How many correct links were retrieved?”
 - macro/micro precision, recall, F-measure
- Ranking-based

Common set-based metrics

- Accuracy

$$A = \frac{|\{\mathcal{C}_{i,0} | \mathcal{C}_{i,0} = \mathcal{G}\}|}{N}$$

- Precision

$$P_{\mathcal{C}} = \frac{|\{\mathcal{C}_i | \mathcal{C}_i \neq \emptyset \wedge \mathcal{G}_i \in \mathcal{C}_i\}|}{|\{\mathcal{C}_i | \mathcal{C}_i \neq \emptyset\}|}$$

- Recall

$$R_{\mathcal{C}} = \frac{|\{\mathcal{C}_i | \mathcal{G}_i \neq \text{NIL} \wedge \mathcal{G}_i \in \mathcal{C}_i\}|}{|\{\mathcal{G}_i | \mathcal{G}_i \neq \text{NIL}\}|}$$

N	Number of queries in data set
\mathcal{G}	Gold standard annotations for data set ($ \mathcal{G} = N$)
\mathcal{G}_i	Gold standard for query i (KB ID or NIL)
\mathcal{C}	Candidate sets from system output ($ \mathcal{C} = N$)
\mathcal{C}_i	Candidate set for query i
$\mathcal{C}_{i,j}$	Candidate at rank j for query i (where $\mathcal{C}_i \neq \emptyset$)

Common ranking-based metrics for entity linking

- Recall @ k
- Precision @ k
- R-precision
- Mean average precision
- Mean reciprocal rank

Test collections

Entity linking test collections

- Wikipedia
- MSNBC
- AQUAINT
- ACE
- Twitter
- AIDA (CoNLL)
- IITB (web data)
- INEX link-the-wiki
- TREC knowledge base acceleration (KBA)
- TAC knowledge base population (KBP)

Wikipedia (for evaluation)

- Widely used
- Pros
 - cheap and easy; the links are already provided
- Cons
 - biased (style guides!)
 - specific scenario
 - unbalanced

MSNBC

[Cucerzan 2007]

- 20 news articles
- Linked to 2006 Wikipedia
 - 756 total links; 127 of these are NIL
- Focus: disambiguate entities after NER and co-reference resolution
 - all mentions of all the detected entities are linked
- Collected by correcting the output of a system

See <http://research.microsoft.com/en-us/um/people/silviu/WebAssistant/TestData/>

AQUAINT

[Milne & Witten 2008]

- 50 news articles
 - 449 links, obtained using Amazon mechanical turk
- subset of AQUAINT newswire, annotated to mimic Wikipedia hyperlink structure
 - only first mentions of “important” titles were linked
 - uninteresting and redundant mentions of the same title not linked

See <http://cogcomp.cs.illinois.edu/page/resources/data>

ACE

[Ratinov et al. 2011]

- Subset of ACE co-reference data set
 - mentions and their types are given
 - co-references resolved
- First nominal mentions of each co-reference chain are linked
 - Amazon mechanical turk
 - accuracy of majority vote ~85%
 - manually corrected

See <http://cogcomp.cs.illinois.edu/page/resources/data>

Twitter

[Meij et al. 2012]

- Tweets taken from “verified accounts,” so relatively clean
- ~500 tweets, manually linked to Wikipedia
 - ~2 entity links per tweet on average

See <http://edgar.meij.pro/dataset-adding-semantics-microblog-posts/>

Task	Name	Year	Source	All Mentions	Instances
CDCR	John Smith	1998	News	✗	197
CDCR	WePS 1	2007	Web	✗	3,489
CDCR	Day et al.	2008	News	✓	3,660
CDCR	WePS 2	2008	Web	✗	3,432
CDCR	WePS 3	2009	Web	✗	31,950
wikify	Mihalcea	2007	Wiki	✓	7,286
wikify	Kulkarni	2009	Web	✓	17,200
wikify	Milne	2010	Wiki	✓	11,000
NEL	Cucerzan	2007	News	✓	797
NEL	TAC 09	2009	News	✗	3,904
NEL	Fader	2009	News	✗	500
NEL	TAC 10	2010	News, Blogs	✗	3,750
NEL	Dredze	2010	News	✗	1,496
NEL	Bentivogli	2010	News, Web, Transcripts	✓	16,851
NEL	Hoffart	2011	News	✓	34,956

Table taken from Hachey et al. (2013). **Evaluating Entity Linking with Wikipedia**. In AI '13.

TAC

[McNamee et al. 2010]

- Target: KB from Wikipedia (~800k instances)
 - infoboxes; article text; type
- Query
 - document ID (news, web, blog)
 - mention string (occurring at least once in that doc)
- Focus on ambiguous mentions
 - collected by cherry-picking ‘interesting’ mentions, rather than systematically annotating all mentions
- Explicit NILs (> 50% of the queries)

	TAC 2009 test		TAC 2010 train		TAC 2010 test	
$ \mathcal{Q} $	3,904		1,500		2,250	
KB	1,675	(43%)	1,074	(72%)	1,020	(45%)
NIL	2,229	(57%)	426	(28%)	1,230	(55%)
PER	627	(16%)	500	(33%)	751	(33%)
ORG	2710	(69%)	500	(33%)	750	(33%)
GPE	567	(15%)	500	(33%)	749	(33%)
News	3904	(100%)	783	(52%)	1500	(67%)
Web	0	(0%)	717	(48%)	750	(33%)
Acronym	827	(21%)	173	(12%)	347	(15%)

$ \mathcal{E} $	560		—		871	
KB	182	(33%)	462	(—)	402	(46%)
NIL	378	(67%)	—	(—)	469	(54%)
PER	136	(24%)	—	(—)	334	(38%)
ORG	364	(65%)	—	(—)	332	(38%)
GPE	60	(11%)	—	(—)	205	(24%)

Table taken from Hachey et al. (2013). **Evaluating Entity Linking with Wikipedia**. In AI '13.

DIY Entity Linking

- Target KB (Wikipedia)
 - wikipedia-miner
 - Google's Dictionaries for Linking Text, Entities and Ideas
- Test collection
- Evaluation metrics

Meta-evaluations

- [Hachey et al. 2013]
- [Cornolti et al. 2013]

Evaluating Entity Linking with Wikipedia

[Hachey et al. 2013]

- Named entity linking, a.k.a., “NEL”
 - include NILs
 - Wikipedia articles not always named entities
- Explicit focus on separating “search” (LG) and “disambiguation” (DA)
- Reimplement and evaluate three NEL systems
 - [Bunescu & Pasça 2006]
 - [Cucerzan 2007]
 - [Varna et al. 2009] (TAC system paper)

System	Extractor	Condition	Searcher							Disambiguator
			Title	Redirect	Link	Truncated	Bold	DABTitle	Filter	
Bunescu and Pașca (2006)	NER	NA	✓	✓				✓	NA	SVM rank over cosine and mention context word×category features
Cucerzan (2007)	NER, coreference expansion	NA	✓	✓	✗	✓		✓	NA	Scalar product between candidate category/term vector and document-level vector
Varma et al. (2009)	NER, acronym expansion	if acronym								Cosine between candidate article term vector and mention context vector
		if expandable	✓							
		else	✓	✓			✓	✓	NA	
		else								
		search 1	✓							
		if no candidates	✓	✓			✓	✓	NA	

Table taken from Hachey et al. (2013). **Evaluating Entity Linking with Wikipedia**. In AI '13.

Alias	Source	$\langle C \rangle$	$P_{\mathcal{C}}^{\infty}$	$R_{\mathcal{C}}^{\infty}$	P_{\emptyset}	R_{\emptyset}
Title		0.2	83.5	37.2	68.1	96.5
Redirect		0.1	74.6	20.0	62.1	96.2
Link		4.2	55.7	80.1	88.6	59.5
Bold		1.6	45.1	48.8	71.7	67.2
Hatnote		0.0	42.6	1.2	57.7	99.9
Truncated		1.2	37.8	24.5	62.2	78.6
DABTitle		3.5	34.2	29.3	58.7	65.1
DABRedirect		2.7	34.0	18.9	57.9	77.3

Table taken from Hachey et al. (2013). **Evaluating Entity Linking with Wikipedia**. In AI '13.

System	A	A_C	A_\emptyset
NIL Baseline	57.1	0.0	100.0
Title Baseline	71.0	37.2	96.5
+ Redirect Baseline	76.3	54.6	92.6
Bunescu and Paşa	77.0	67.8	83.8
Cucerzan	78.3	71.3	83.5
Varma et al. Replicated	80.1	72.3	86.0
TAC 09 Median	71.1	63.5	78.9
TAC 09 Max (Varma)	82.2	76.5	86.4

Table taken from Hachey et al. (2013). **Evaluating Entity Linking with Wikipedia**. In AI '13.

A Framework for Benchmarking Entity-Annotation Systems

[Cornolti et al. 2013]

- Compare five publicly available entity linkers
 - [Hoffart et al. 2007] (AIDA)
 - [Ratinov et al. 2011]
 - [Ferragina & Scaiella 2010] (TAGME)
 - [Milne & Witten 2008] (wikipedia-miner)
 - DBpedia Spotlight
- And also investigate parameter/cut-off settings

See <http://acube.di.unipi.it/> (not available yet).

A Framework for Benchmarking Entity-Annotation Systems

[Cornolti et al. 2013]

- On five publicly available test collections
 - AIDA **[Hoffart et al. 2007]**
 - based on CoNLL 2003: noun annotations
 - 1393 Reuters newswire articles
 - hand-annotated all nouns with entities in YAGO2
 - AQUAINT **[Milne & Witten 2008]**
 - MSNBC **[Cucerzan 2007]**
 - IITB **[Kulkarni et al. 2010]** (web data)
 - Twitter **[Meij et al. 2012]**

A Framework for Benchmarking Entity-Annotation Systems

[Cornolti et al. 2013]

- Main findings:

A Framework for Benchmarking Entity-Annotation Systems

[Cornolti et al. 2013]

- Main findings:
 - Come see for yourself
 - Friday, 17:00
 - Practice and Experience Track - System / Infrastructure
 - Room Itamaraty

Open challenges

Open challenges

- Difficulty prediction
 - similar to ambiguity, but not the same
 - dependent on context, candidate links, ...
- Cross-lingual entity linking [**Wang et al. 2013**]
- Cross-KB entity linking (“Freebase”)
 - use Wikipedia as pivot
 - directly
 - lexical matching
 - machine learning (if annotators/training data available)

Learning/Updating the KB

- Parallel, continuous streams of items
 - news
 - tweets, blogs, status updates
 - queries, clicks
 - web pages, RDFa/schema.org
 - etc.
- Given an entity
 - “What is new?” What do I need to know now?”
 - Add: personal
 - Add: social



twitter



Learning/Updating the KB

- Why?
 - going beyond ten blue links
 - dynamic (entity) overview pages
 - dynamic entity timelines
 - push interesting updates to the user
 - predict/Explain (non-trivial) events
- And: feed back novel information into the knowledge base(s)
 - create new nodes
 - create new links

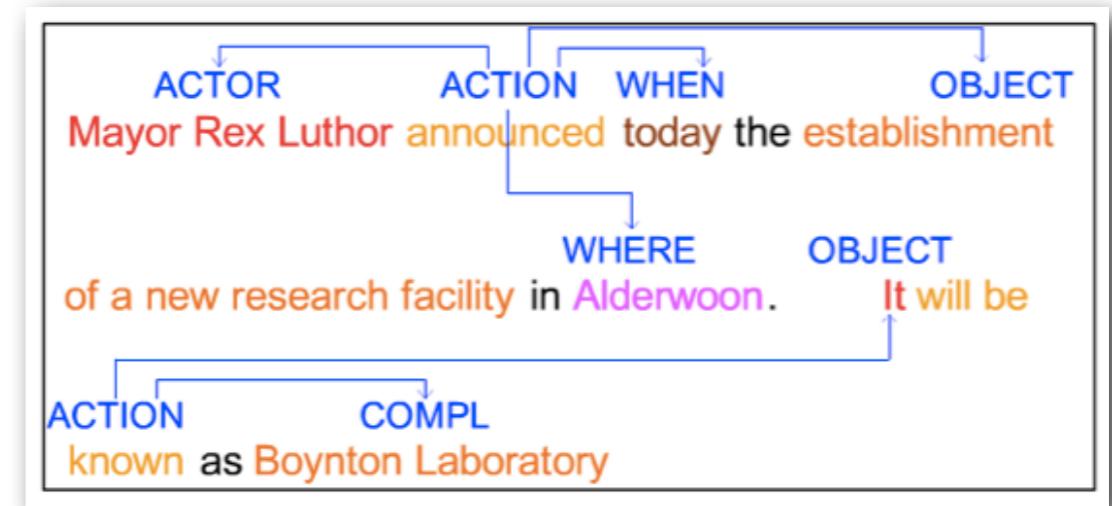


twitter



Learning/Updating the KB: ingredients?

- Accurate entity linking
 - real-time
 - cross-item
 - cross-genre
 - cross-vertical
- Detect bursts, events
- What is being said?
 - aspects, attributes, relations
- Correlate with already known facts



Open challenges

- Generic test collections
 - What's the task?
 - TAC KBP?
 - entity link ranking?
 - known-item finding, i.e., top-1 retrieval?
- Moving beyond entities
 - events/news
 - concepts
 - relations
 - “things in the world”

Open challenges

- What if there is no/little textual evidence?
- Moving beyond "ad hoc" entity linking:
incorporate contextual evidence in the task
(and evaluation)
 - users
 - history
 - profile
 - social
 - trending

Follow-up reading

- Detecting unlinkable entities [Lin et al. 2012a]
- Linking entities to any database [Sil et al. 2012]
- Automatically generating Wikipedia articles: A structure-aware approach
- Scaling up to the web [Lin et al. 2012b]
- Serendipitous suggestions based on personalized entity links [Bordino et al. 2013]

References

<http://www.mendeley.com/groups/3339761/entity-linking-and-retrieval-tutorial-at-www-2013-and-sigir-2013/papers/added/0/tag/entity+linking/>

References

The screenshot shows a Mendeley group page titled "Entity Linking and Retrieval – Tutorial at WWW 2013 and SIGIR 2013". The page displays three research papers:

- Analysis and Enhancement of Wikification for Microblogs with Context Expansion.** By Taylor Cassidy, Heng Ji, Lev-Arie Ratinov, Arkaitz Zubiaga, Hongzhao Huang. Published in COLING 2012 (2012).

Disambiguation to Wikipedia (D2W) is the task of linking mentions of concepts in text to their corresponding Wikipedia entries. Most previous work has focused on linking terms in formal texts (e.g. newswire) to Wikipedia. Linking terms in short...

Added 1 minute ago | 1 reader
- Microblog-genre noise and impact on semantic annotation accuracy** by Leon Derczynski, Diana Maynard, Niraj Aswani, Kalina Bontcheva in HT 2013 (2013).

Using semantic technologies for mining and intelligent information access to microblogs is a challenging, emerging research area. Unlike carefully authored news text and other longer content, tweets pose a number of new challenges, due to their...

Added 11 minutes ago
- Entity Disambiguation with Freebase** by Zhicheng Zheng, Xiancse Si, Fangtao Li, Edward Y. Chang, Xiaoyan Zhu in WI-IAT.

The right sidebar shows "Top tags in this group" including entity linking, Wikipedia, TAC, commonness, SVM, graph, relatedness, naive bayes, pagerank, keyphraseness, Twitter, centrality, meta evaluation, NER, word sense disambiguation, random forests, Freebase, tagme, local, and web.

<http://www.mendeley.com/groups/3339761/entity-linking-and-retrieval-tutorial-at-www-2013-and-sigir-2013/papers/added/0/tag/entity+linking/>

Hands-on entity linking