

Program - morning

| | |
|----------------------|----------------------------------|
| 08:30 - 08:45 | Welcome, introduction |
| 08:45 - 10:30 | Entity linking - methods |
| 10:30 - 11:00 | Coffee break |
| 11:00 - 11:15 | Entity linking - QA |
| 11:15 - 12:15 | Entity linking - hands-on |
| 12:15 - 13:30 | Lunch |

Outline

- **Background**
 - AQUAINT dataset
 - DBpedia Spotlight
 - Wikipedia Miner
 - Wikipedia-based measures
- **Exercises**
 - Entity Linking using Webservice
 - Evaluation Entity Linking
 - Lexical Matching on Anchor Text

AQUAINT

[Milne & Witten 2008]

Antique
Collectible
charge_card
Online_auction_business_model
Beanie_Baby
Christie's
Auction
Sports_memorabilia
New_York_City
Sotheby's
Baseball
Rembrandt

Sotheby's Announces Internet Site 20-10-1999, NEW YORK (AP)

Home shoppers can buy airline tickets, Beanie Babies and garden tools via computer. Why not a \$9 million Rembrandt painting? Sotheby 's, the New York auction house , said Tuesday that it will spend \$25 million to launch sothebys.com, an Internet site where the public will be able to bid on antiques , art, jewelry, sports memorabilia and other collectibles.

The value of items on the virtual auction block initially will be capped at about \$10,000. More expensive items will be posted when imaging technology improves enough to give clear views of items from multiple angles. "There's no point in having a fuzzy image on a \$100,000 item," said Susan Solomon, head of the new venture. Sotheby's first major online auction, planned for July, will feature baseball memorabilia.

While online auction sites have become extremely popular, consumer groups say the sites have produced more consumer complaints than any other Internet category because of shoddy merchandise and other fraud. Sotheby's said it would only sell merchandise from the auction house itself or reputable art dealers. And bids will only be accepted from customers who have registered their charge card numbers in advance.

Officials at Christie 's, another leading New York auction house, said they were also considering an online auction site.



Confidence:

0.0

Language: English



Annotation score:

0.0

☐ n-best candidates

SELECT TYPES...

ANNOTATE

First documented in the 13th century, [Berlin](#) was the capital of the Kingdom of [Prussia](#) (1701–1918), the [German Empire](#) (1871–1918), the [Weimar Republic](#) (1919–33) and the [Third Reich](#) (1933–45). [Berlin](#) in the 1920s was the third largest [municipality](#) in the world. After [World War II](#), the city became divided into [East Berlin](#) – the capital of [East Germany](#) – and [West Berlin](#), a [West German exclave](#) surrounded by the [Berlin Wall](#) from 1961–89. Following [German reunification](#) in 1990, the city regained its status as the capital of [Germany](#), hosting 147 foreign embassies.

BACK TO TEXT

<http://dbpedia-spotlight.github.io/demo/>

http://spotlight.dbpedia.org/rest/annotate

```
01 {
02   "@text": "President Obama called Wednesday on Congress to extend a tax break for students
03   included in last year's economic stimulus package, arguing that the policy provides more
04   generous assistance.",
05   "@confidence": "0.1",
06   "@support": "10",
07   "@types": "",
08   "@sparql": "",
09   "@policy": "whi",
10   "Resources": {
11     "@URI": "http://dbpedia.org/resource/Presidency_of_Barack_Obama",
12     "@support": "134",
13     "@types": "DBpedia:OfficeHolder,DBpedia:Person,Schema:Person,Freebase:/book/
14     book_subject,Freebase:/book,Freebase:/book/periodical_subject,Freebase:/media_common/
15     quotation_subject,Freebase:/media_common,DBpedia:TopicalConcept",
16     "@surfaceForm": "President Obama",
17     "@offset": "0",
18     "@similarityScore": "0.18051494657993317",
19     "@percentageOfSecondRank": "-1.0"
20   },
21   {
22     "@URI": "http://dbpedia.org/resource/Fifty_State_Initiative",
23     "@support": "117",
24     "@types": "Freebase:/fictional_universe/fictional_organization,Freebase:/
25     fictional_universe",
26     "@surfaceForm": "called",
27     "@offset": "16",
28     "@similarityScore": "0.03411872312426567",
29     "@percentageOfSecondRank": "-1.0"
30   }
31 }
```

Filter settings

of inlinks

Entity URI

Cosine Similarity



wikipedia**miner**

[home](#) [demos](#) [services](#) [help](#)

introduction

the demos

search

compare

annotate

Text to annotate

Wikipedia is a free, multilingual encyclopedia project supported by the non-profit Wikimedia Foundation. Its name is a portmanteau of the words wiki (a technology for creating collaborative websites, from the Hawaiian word wiki, meaning 'fast') and encyclopedia.

Wikipedia's 12 million articles (2.77 million in English) have been written collaboratively by volunteers around the world, and almost all of its articles can be edited by anyone who can access the Wikipedia website. Launched in January 2001 by Jimmy Wales and Larry Sanger, it is currently the most popular general reference work on the Internet.

[show options](#)

Annotate

Annotated text

MediaWiki Markup

Detected Topics

[[Wikipedia]] is a free, multilingual encyclopedia project supported by the [[Nonprofit organization|non-profit]] [[Wikimedia Foundation]]. Its name is a [[portmanteau]] of the words [[wiki]] (a technology for creating collaborative websites, from the [[Hawaiian language|Hawaiian word]] wiki, meaning 'fast') and

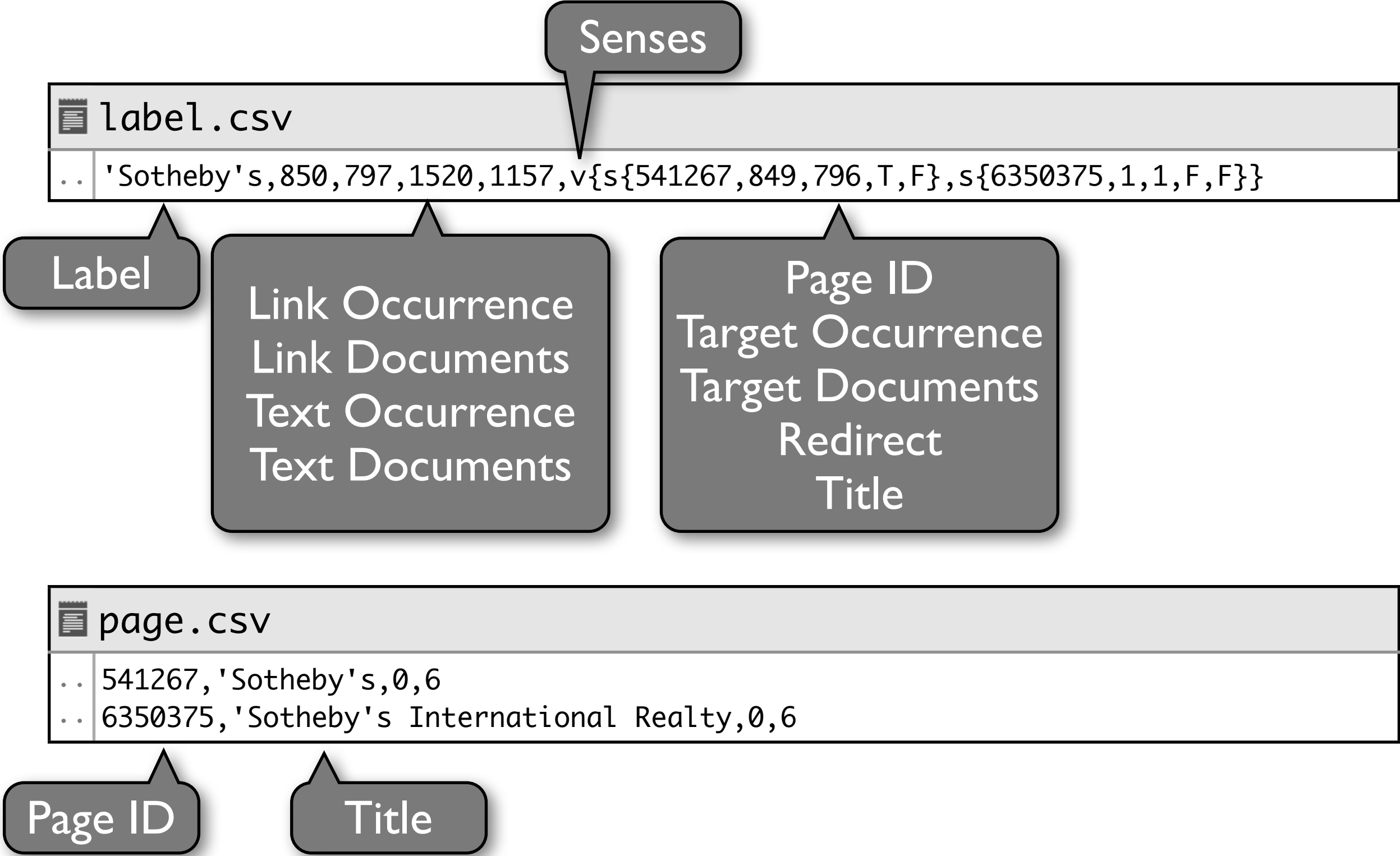
Wikipedia's 12 mill been written collab and almost all of its articles can be edited by anyone who can access the Wikipedia website. Launched in January 2001 by [[Jimmy Wales]] and [[Larry Sanger]], it is currently the most popular general reference work on the Internet.

no definition available

59% probability of being a link

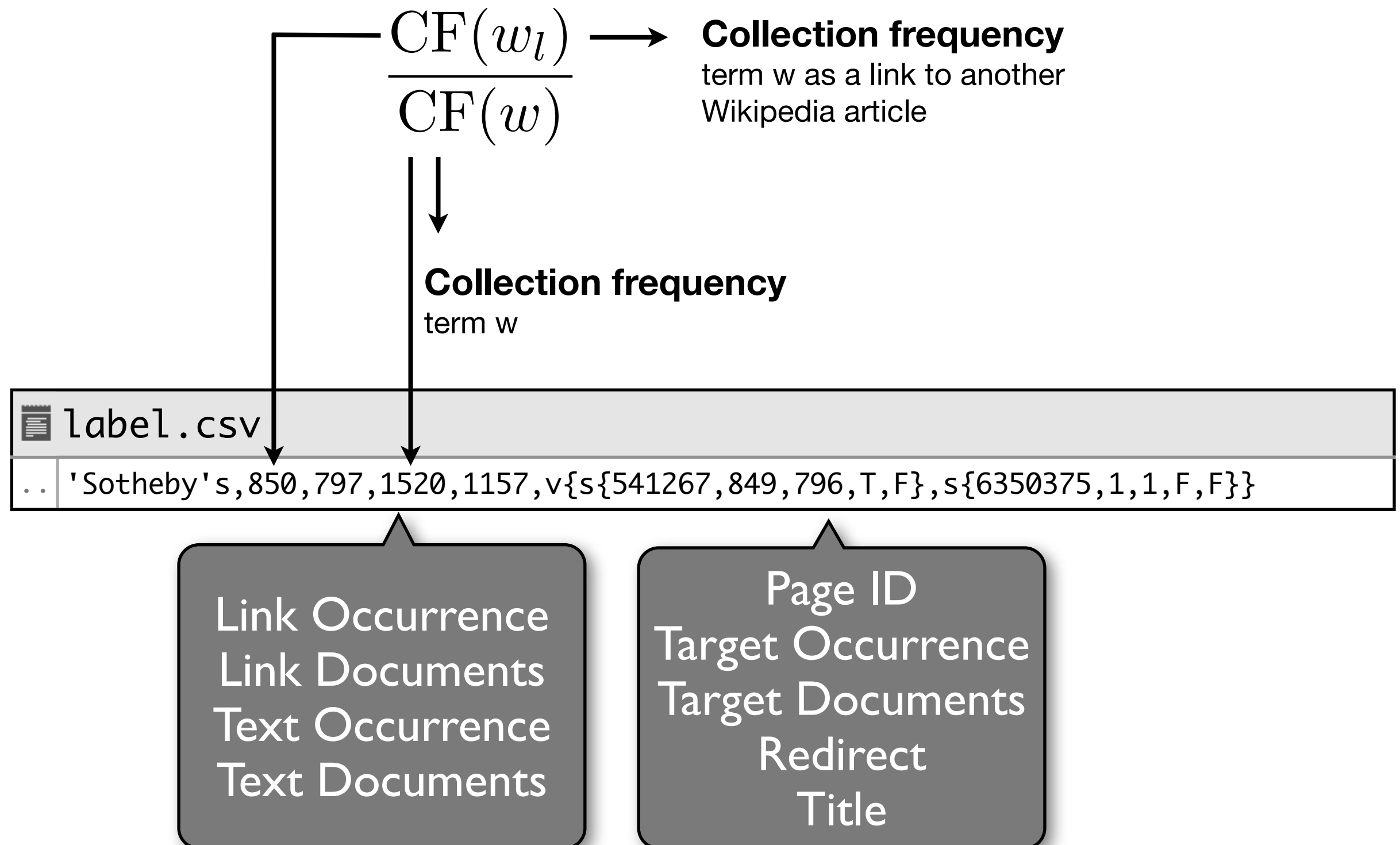
Wikipedia Miner: CSV Summary Files

[Milne & Witten 2008b]



Wikipedia-based measures

Keyphraseness [Mihalcea & Csomai 2007]



Wikipedia-based measures

Commonness [Medelyan et al. 2008]

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

Number of links

with target c' and anchor text w

| |
|------------------------------------------------------------------------------|
| label.csv |
| .. 'Sotheby's',850,797,1520,1157,v{s{541267,849,796,T,F},s{6350375,1,1,F,F}} |

Link Occurrence
Link Documents
Text Occurrence
Text Documents

Page ID
Target Occurrence
Target Documents
Redirect
Title

<http://bit.ly/ELR-course>

Codecademy

Learn

Teach

Entity Linking and Retrieval

Course written by [Daan Odijk](#)

This course is created as a tutorial for Entity Linking and Retrieval. It covers web services, evaluation and fundamental methods. This Codecademy course is part of a tutorial created for WWW2013 and SIGIR2013, see: <http://bit.ly/ELR-slides>.

1. Entity Linking using Web Services

In this first section, we'll create a page where we can enter a text and then perform entity linking on this text. For this, we will use the DBpedia Spotlight web service.

2. Evaluating Entity Linking

In this section we will evaluate an entity linking webservice using a standard dataset. We will evaluate the DBpedia Spotlight Webservice on the AQUAINT dataset.

3. Entity Linking: Lexical matching of anchor text

In this exercise we will build the basic matching for lexical matching of anchor text from the ground up. This exercise is based on the University of Amsterdam Semanticizer and the University of Waikato Wikipedia Miner.

4. Entity Retrieval

In this section we will evaluate Entity Retrieval using Freebase.