ICSC 2011

Stanford University
Palo Alto, California

September 19-21, 2011

WiGipedia

Improving Structured Data in Wikipedia

S. Bostandjiev, J. O'Donovan, C. Hall, B. Gretarsson, T. Hollerer
University of California, Santa Barbara

http://wigipedia-online.com

UCSB

UNIVERSITY OF CALIFORNIA · SANTA BARBARA

WiGipedia: Motivation

What is wrong with data spread across multiple Wiki articles?

Chevrolet

Type	Division of GM
Industry	Automotive
Founded	November 3, 1911
Founder(s)	Louis Chevrolet, William C. Durant
Headquarters	Detroit, Michigan, U.S.
Products	Automobiles
Parent	General Motors Company
Website	Chevrolet.com

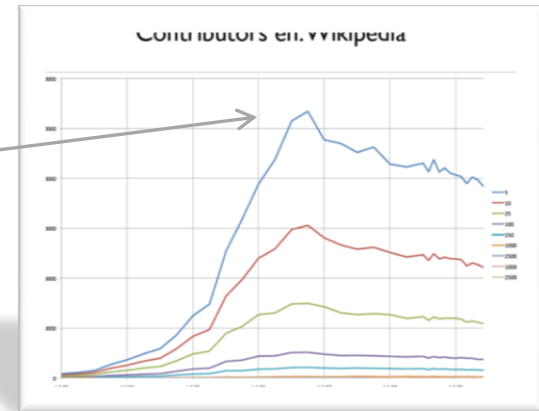Cadillac Motor Car Division

Type	Division of GM
Industry	Automobiles
Founded	August 22, 1902
Headquarters	Warren, Michigan, U.S.[1]
Key people	Henry M. Leland, founder Henry Ford, original founder (Henry Ford Company)
Products	Luxury vehicles
Parent	General Motors Company
Website	cadillac.com

WiGipedia: Motivation

Inconsistencies across structured data in Wikipedia

WiGipedia: **Motivation**

Decline in Wiki editing
(since 2006)

- – Increasingly complex security
  policies

- – Saturation of simpler articles

- – Wiki markup is too complicated
  for the average user


```
{{Other uses}}
{{Use dmy dates|date=July 2011}}
{{Cleanup-rewrite|date=May 2010}}
{{Refimprove|date=April 2009}}
{{Infobox company
|logo = [[File:Caddynew.png|175px]]
|company_name = Cadillac Motor Car Division
|company_type = [[Division
(organization)|Division]] of GM
|foundation = August 22, 1902
|location_city = [[Warren, Michigan|Warren]],
Michigan, U.S.<ref>{{cite web|title=Cadillac
Centennial Anniversary
(brochure)|url=http://www.motorcities.org/pdf/yrt_t
ours_cent_cadillac.  pdf}}</ref>
```

(Andrew Lih, Wikimedia Foundation)

WiGipedia: Motivation

Semantic Links

Encourage people to **create** new semantic links in Wikipedia

Ensure the links are **consistent** across existing data

Visual Wikipedia Editing

Provide **visual editing** as an alternative to the wiki markup (for structured data)

Context

Provide context to
- single wiki article
- set of wiki articles

Wikipedia → DBpedia → WiGipedia

Wikipedia
structured data
(infoboxes, categories, etc.)

DBpedia
structured semantic data (RDF)
allows SQL-like querying

WiGipedia
visual interface
near single-click Wiki edits

(closing the loop)

Demo

Live demo

http://wigipedia-online.com

How to generate a seed node set?
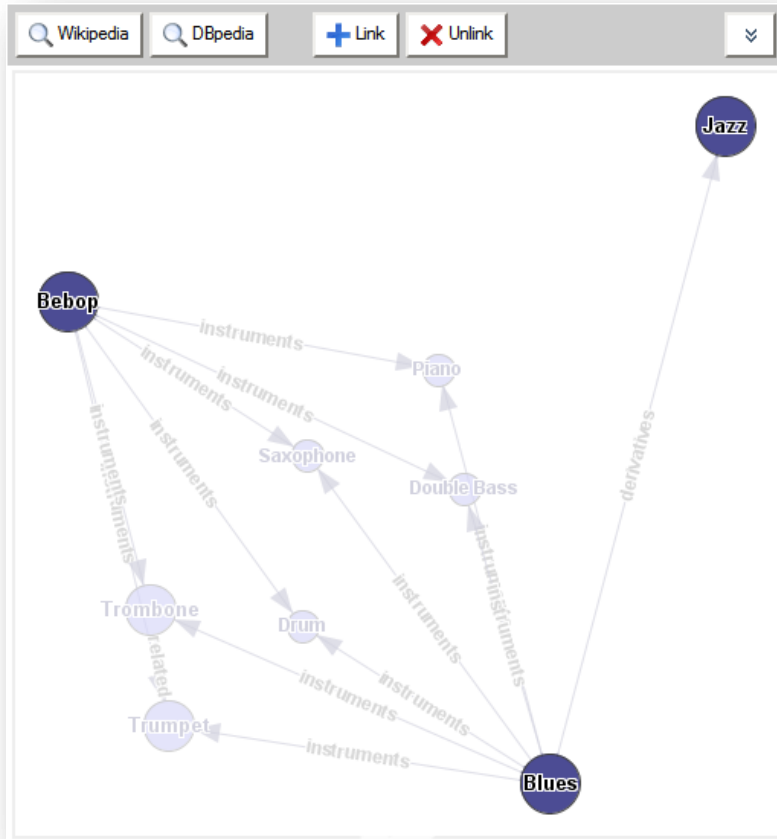
Hits	Article	Hits	Article
26	United_States	6	New_Orleans
11	Bebop	6	Louis_Armstrong
9	Free_jazz	6	Herbie_Hancock
9	Hard_bop	5	Ragtime
9	Miles_Davis	5	Saxophone
8	Blues	5	Rhythm_and_blues
8	Jazz_fusion	5	Avant-garde_jazz
8	Dixieland	5	Big_band
6	Swing_music	5	Modal_jazz
6	Africa	5	Bossa_nova

Table 1: Example collection of the 20 most referenced pages (with occurrence counts) as scraped from the Jazz article on Wikipedia. Note their diversity of 'data types' - places, genres, persons, instruments, and objects.

How to display the data?
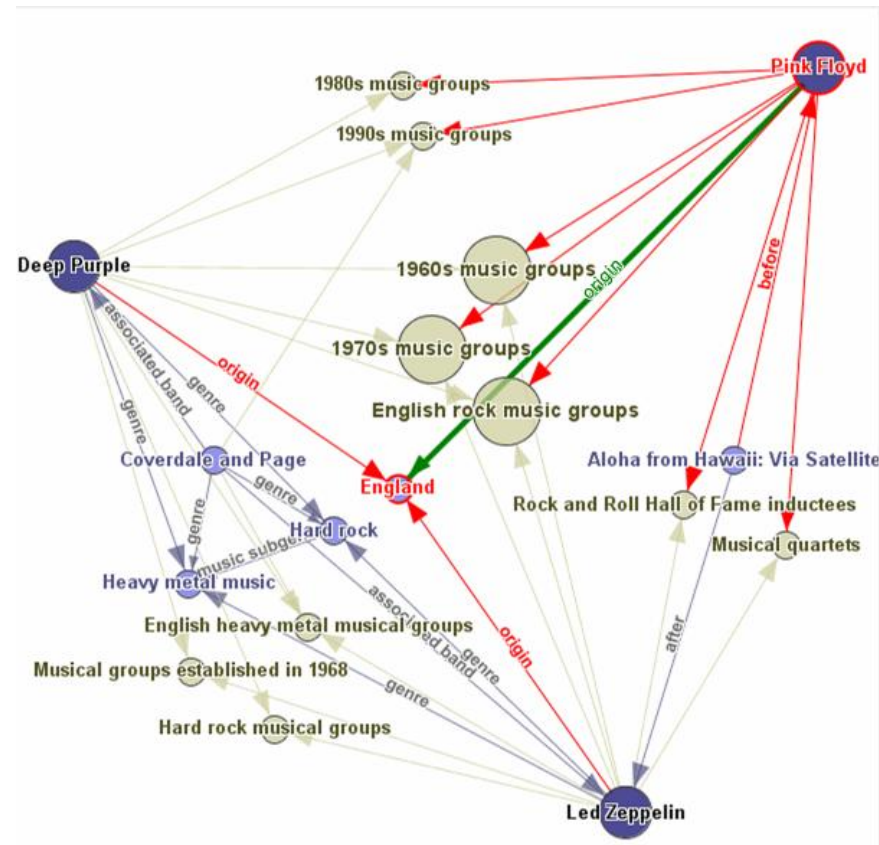
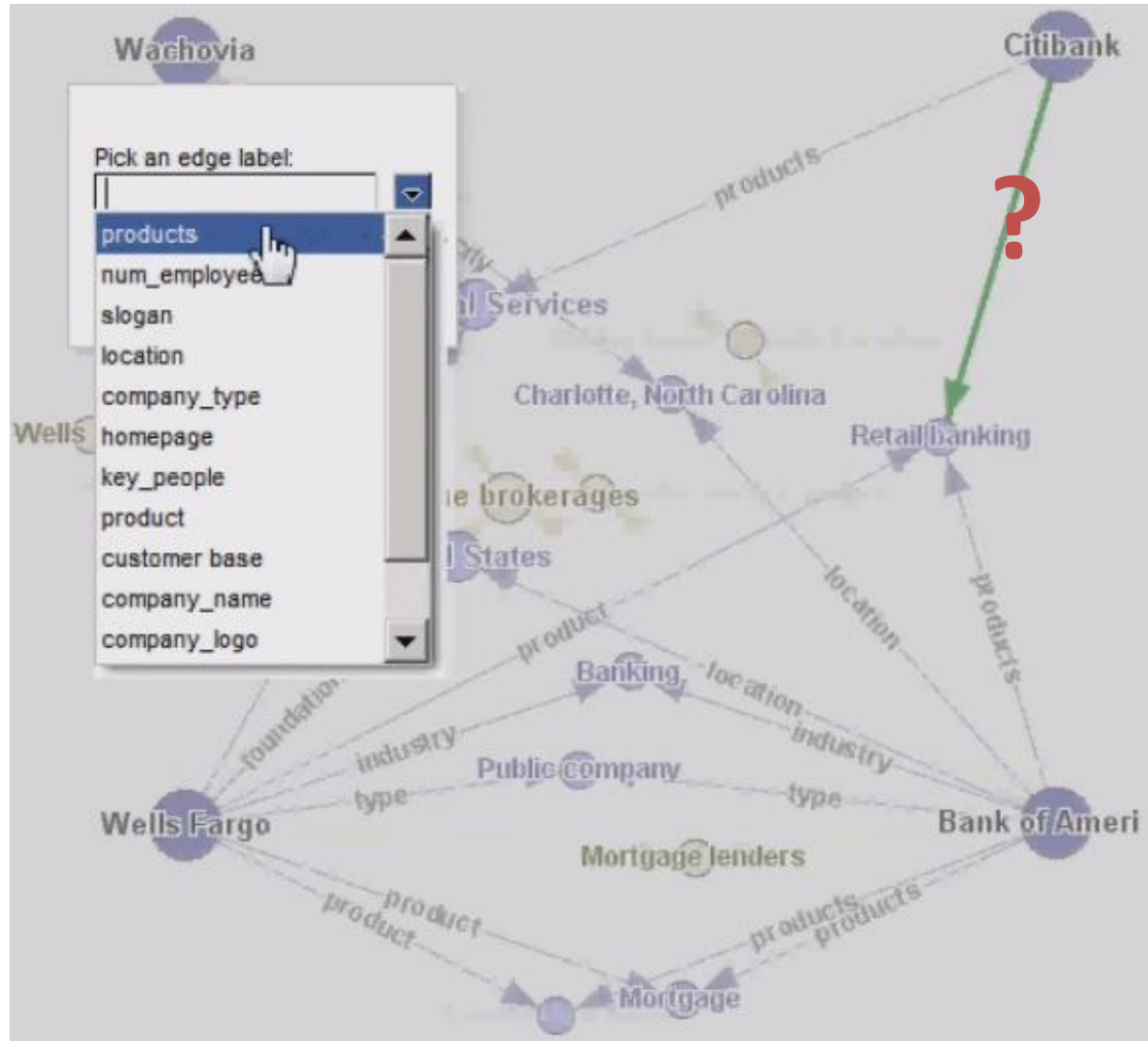|	● Led Zeppelin	● Deep Purple	● Pink Floyd
◉ Aloha from Hawaii: Via Satellite	after ⬆		before ⬆
◉ Coverdale and Page	associated band ⬆	associated band ⬆	
◉ England	origin ↵	origin ↵	origin ↵
◉ Hard rock	genre ↵	genre ↵	
◉ Heavy metal music	genre ↵	genre ↵	
◉ 1960s music groups	✓	✓	✓
◉ 1970s music groups	✓	✓	✓
◉ 1980s music groups	⊘	✓	✓
◉ 1990s music groups	⊘	✓	✓
◉ English heavy metal musical groups	✓	✓	⊘
◉ English rock music groups	✓	✓	✓
◉ Hard rock musical groups	✓	✓	⊘
◉ Musical groups established in 1968	✓	✓	⊘
◉ Musical quartets	✓	⊘	✓
◉ Rock and Roll Hall of Fame inductees	✓	⊘	✓

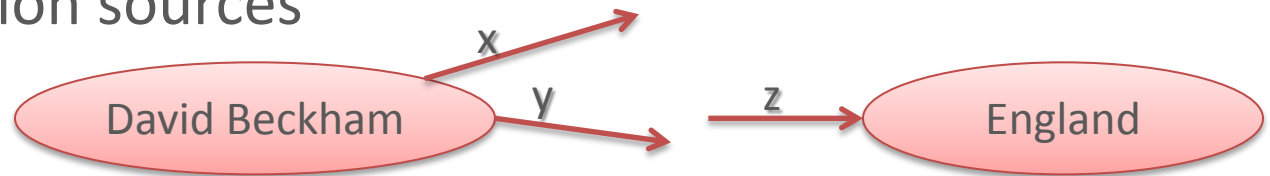● Main Wikipedia articles, ● Other Wikipedia articles, ◉ Wikipedia categories
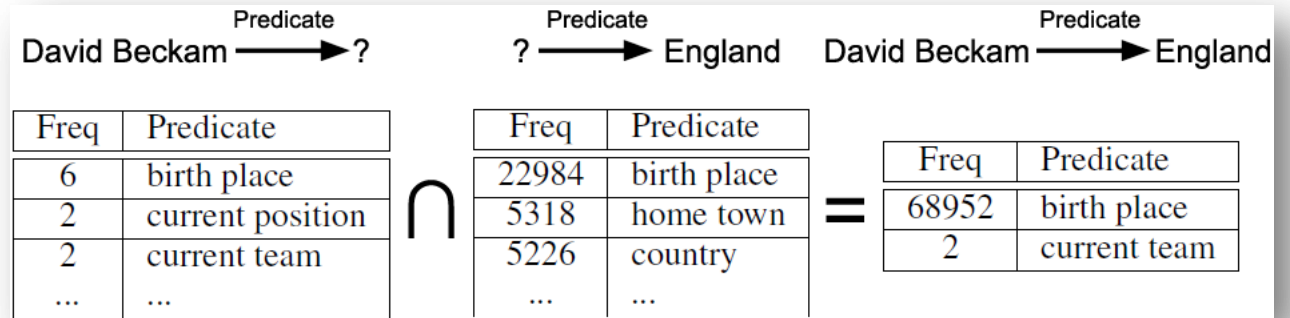
How to label a new link?

How to label a new link?

Recommendation sources

Graph

David Beckham — x, y — z — England

DBpedia

	Predicate			Predicate			Predicate	
David Beckam → ?			? → England			David Beckam → England		

Freq	Predicate
6	birth place
2	current position
2	current team
...	...

∩

Freq	Predicate
22984	birth place
5318	home town
5226	country
...	...

=

Freq	Predicate
68952	birth place
2	current team

Wikipedia Templates

```
{{Infobox football biography
| name           =
| image          =
| caption        =
| fullname       =
| birth_date     =
| birth_place    =
| height         =
| position       =
| currentclub    =
| clubnumber     =
| youthyears1    =
| youthclubs1    =
| collegeyears1  =
```
```
| collegeclubs1       =
| collegecaps1        =
| collegegoals1       =
| years1              =
| clubs1              =
| caps1               =
| goals1              =
| nationalyears1      =
| nationalteam1       =
| nationalcaps1       =
| nationalgoals1      =
| medaltemplates      =
| club-update         =
| nationalteam-update =
}}
```
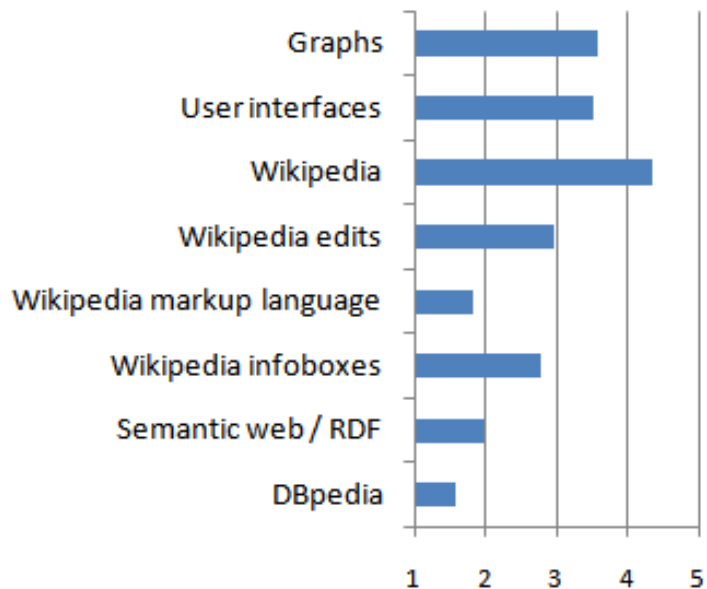
Ranking

User Study

25 participants

How familiar are you with...

Graphs	
User interfaces	
Wikipedia	
Wikipedia edits	
Wikipedia markup language	
Wikipedia infoboxes	
Semantic web / RDF	
DBpedia	

1 2 3 4 5

Tasks:

Pick a few related articles
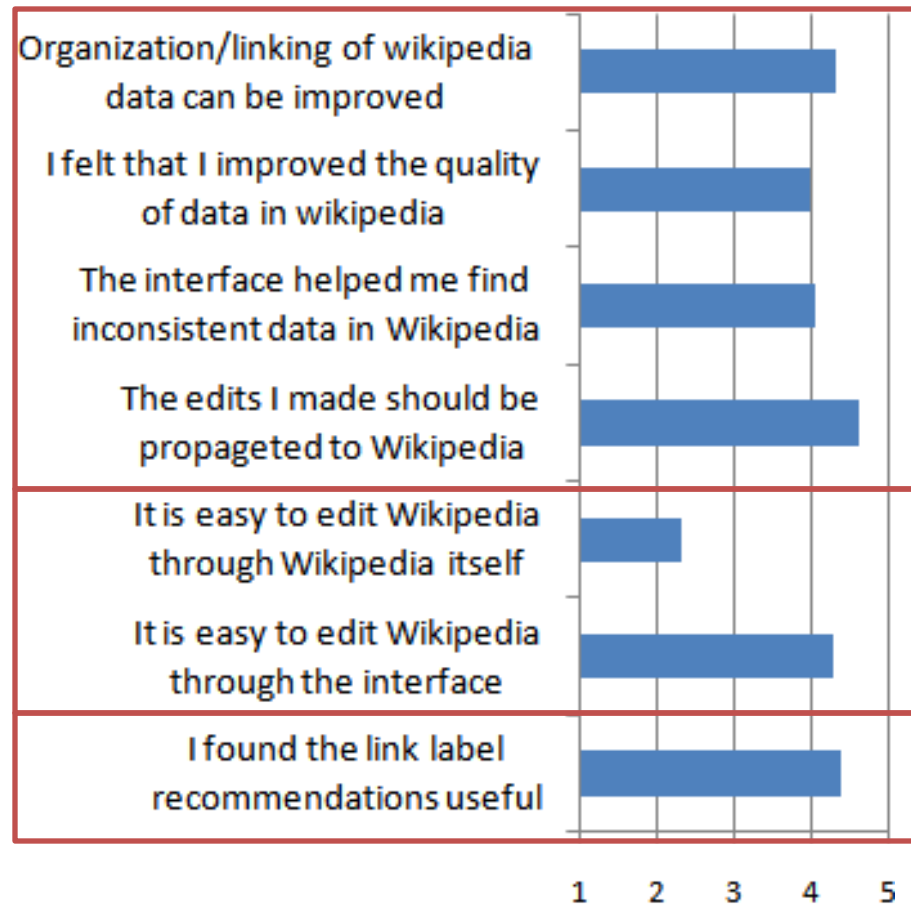
Look at how they are related

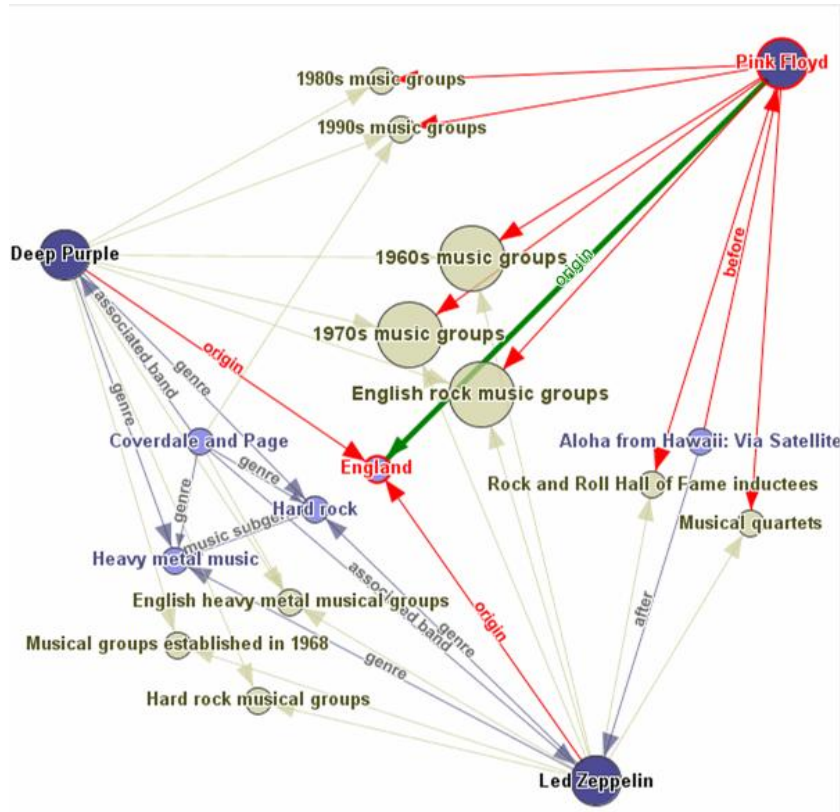**Edit** the links as you see fit:

- Through WiGipedia
- Through Wikipedia

Create a new link and **label** it

Use either **graph or table** or both

User Study

User Study: Graph vs. Table

"It is more **satisfying** to connect things like a web. I could see myself building it, creating something, making the connection."

"The table was more **efficient** and intuitive. It let me look at and compare things side by side."

WiGipedia: Limitations

Works with Structured Wikipedia Data Only

Infoboxes and categories, but not text

Automating Wikipedia Edits

Not always straightforward

Reliance on 3rd Party Data Providers

Correctness of DBpedia crawling & parsing

SPARQL access point speed (generating a graph may take a minute)

WiGipedia: Contributions

Provides Context

Single article: Relevant contextual information per article

Multiple articles: Reveals connections between multiple articles

Reduces Incompleteness

Mechanism to detect and create missing links

Increases Consistency

Promotes consistency and accuracy of structured data spread across multiple articles

Facilitates Editing

Provides visual editing as an alternative to the Wikipedia markup language

Live demo
http://wigipedia-online.com

Contact

Underlying visualization framework
http://wigis.net

alex@cs.ucsb.edu
http://bostandjiev.com