



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://wikipedia-online.com>



WiGipedia: Motivation

What is wrong with data spread across multiple Wiki articles?

Chevrolet



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

Chevrolet



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

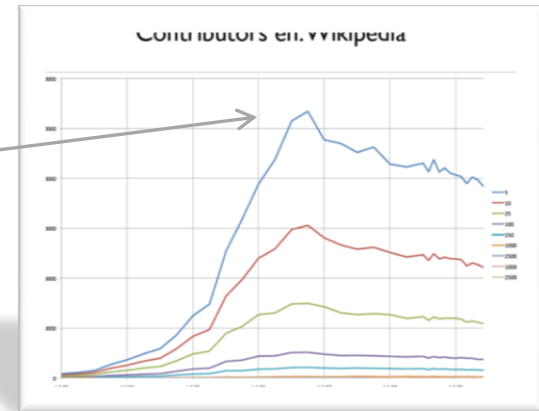


Type	Division of GM
Industry	Automobiles
Founded	August 22, 1902
Headquarters	Warren, Michigan, U.S. ^[1]
Key people	Henry M. Leland, founder Henry Ford, original founder (Henry Ford Company)
Products	Luxury vehicles
Parent	General Motors Company
Website	cadillac.com

WiGikipedia: 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)

WiGikipedia: 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

The screenshot shows the Wikipedia article for Pink Floyd. A semantic network is overlaid on the article, showing connections between Pink Floyd and other entities like Deep Purple, Led Zeppelin, and various music genres. A sidebar table provides background information and former members.

Background information	
Origin	London
Genres	Progressive rock Psychedelic rock
Years active	1965–1996, 2005
Labels	EMI, Harvest, Capitol, Tower, Columbia
Associated acts	Sigma 6, Joker's Wild, Zee
Website	www.pinkfloyd.co.uk www.pinkfloyd.com

Former members	
Syd Barrett	
David Gilmour	
Bob Klose	
Nick Mason	
Roger Waters	
Richard Wright	

Categories: Pink Floyd | Musical groups established in 1965 | Musical groups disestablished in 1996 | English progressive rock groups | Psychedelic rock music groups | 1960s music groups | 1970s music groups | 1980s music groups | 1990s music groups | Musical quartets | Rock and Roll Hall of Fame inductees | Grammy Award winners | Capitol Records artists | Polar Music Prize laureates

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)

Led Zeppelin



Led Zeppelin in 1978. From left to right: John Bonham, Robert Plant, Jimmy Page, John Paul Jones

Background information

Origin London, England

Genres Hard rock, heavy metal, blues-rock, folk rock

Years active 1968-present
< huge >(Reunions: 1985, 1988, 1995, 2007)

Labels Atlantic, Swan Song

Associated acts Page and Plant, The Honeydrippers, The Yardbirds, Them Crooked Vultures

Website ledzeppelin.com

Former members

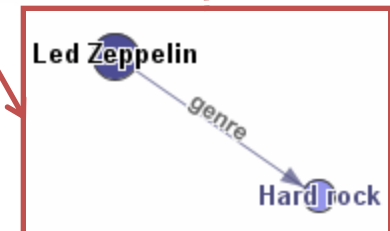
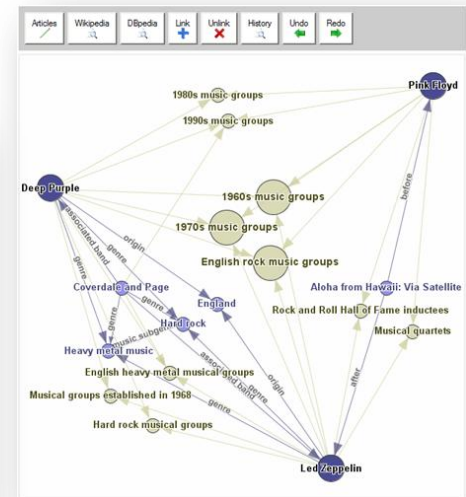
- Jimmy Page
- John Paul Jones
- Robert Plant
- John Bonham

http://dbpedia.org/page/Led_Zeppelin

About: [Led Zeppelin](#)
An Entity in Data Space: dbpedia.org

Led Zeppelin were an English rock band formed in 1968 by Jimmy Page, Robert Plant, John Paul Jones and John Bonham. They were one of the most successful heavy metal bands, helping to pioneer the genre. However, the band's individualist

Property	Value
dbpedia-owl:Artist/genre	<ul style="list-style-type: none"> dbpedia:Hard_rock dbpedia:Folk_rock dbpedia:Blues-rock dbpedia:Heavy_metal_music
dbpedia-owl:Artist/label	<ul style="list-style-type: none"> dbpedia:Swan_Song_Records dbpedia:Atlantic_Records
dbpedia-owl:MusicalArtist/associatedBand	<ul style="list-style-type: none"> dbpedia:The_Yardbirds dbpedia:The_Honeydrippers dbpedia:Page_and_Plant dbpedia:Raising_Sand dbpedia:Them_Crooked_Vultures dbpedia:Band_of_Joy dbpedia:The_Firm_(band) dbpedia:Coverdale_&_Page

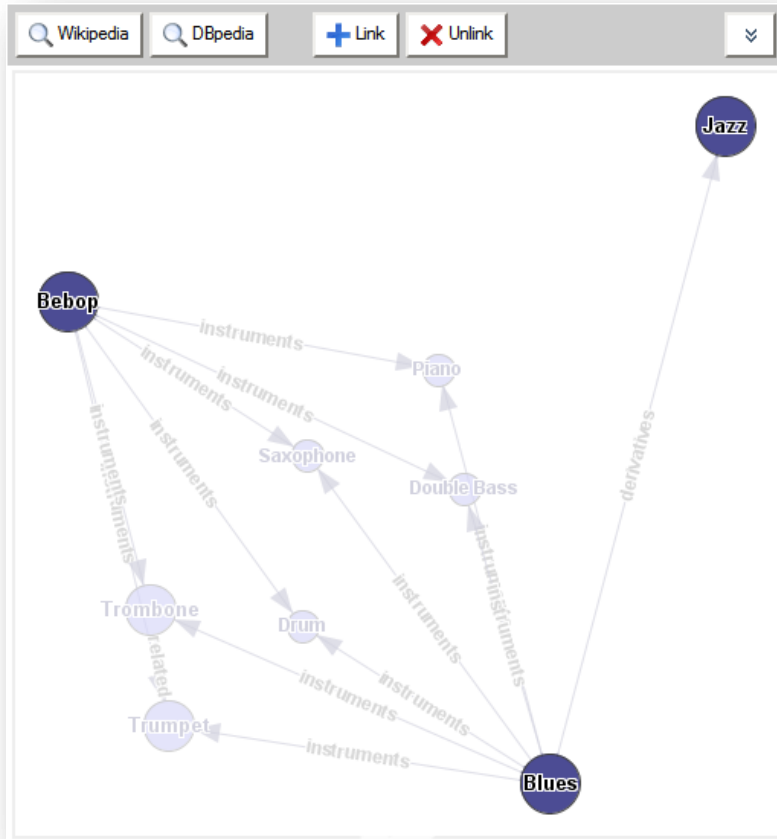


Demo

Live demo

<http://wikipedi-online.com>

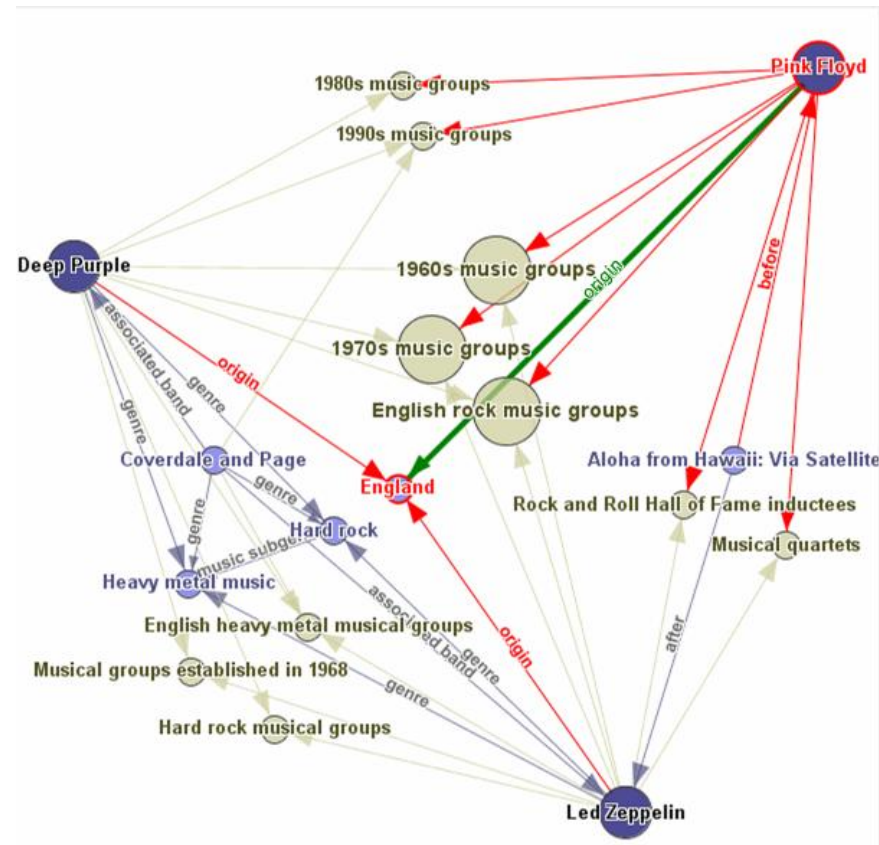
How to generate a seed node set?



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.

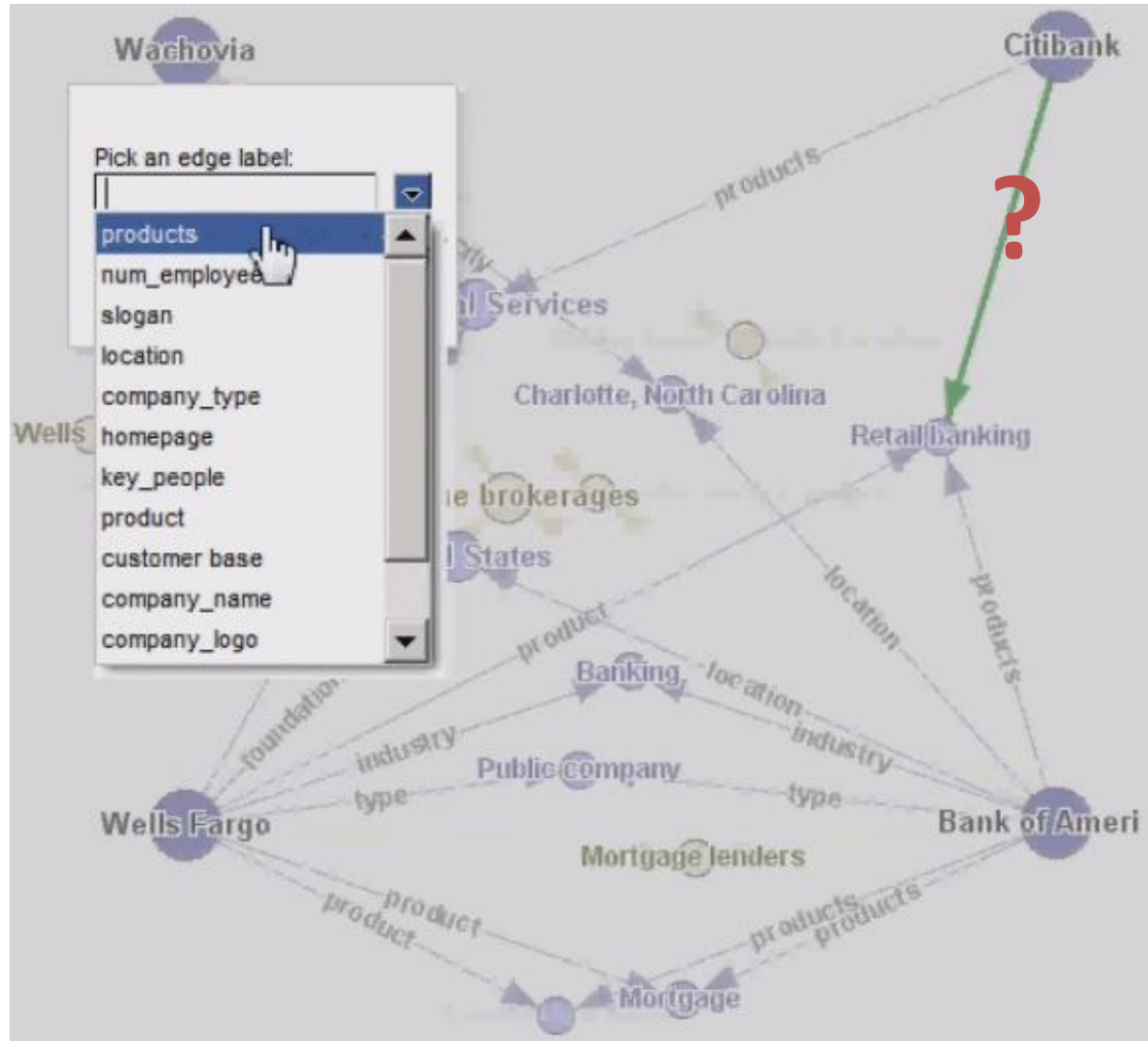
How to display the data?



	● 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

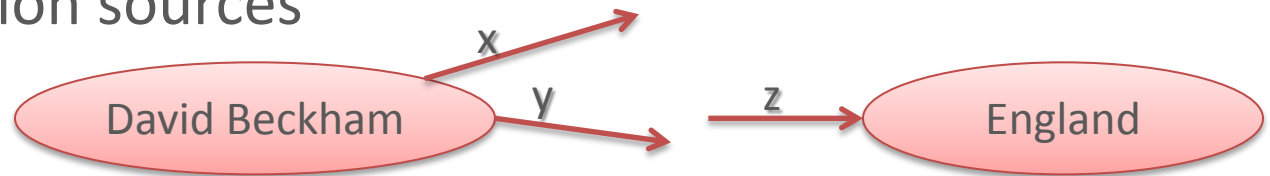
How to label a new link?



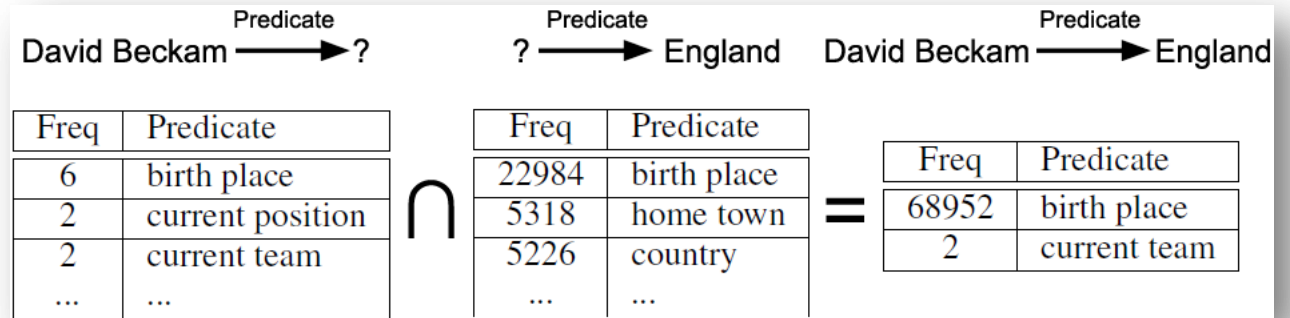
How to label a new link?

Recommendation sources

Graph



DBpedia



Wikipedia Templates

Ranking

```

{{Infobox football biography
| name           =
| image          =
| caption        =
| fullname       =
| birth_date     =
| birth_place    =
| height         =
| position       =
| currentclub    =
| clubnumber     =
| youthyears1   =
| youthclubs1    =
| collegeyears1 =

```

```

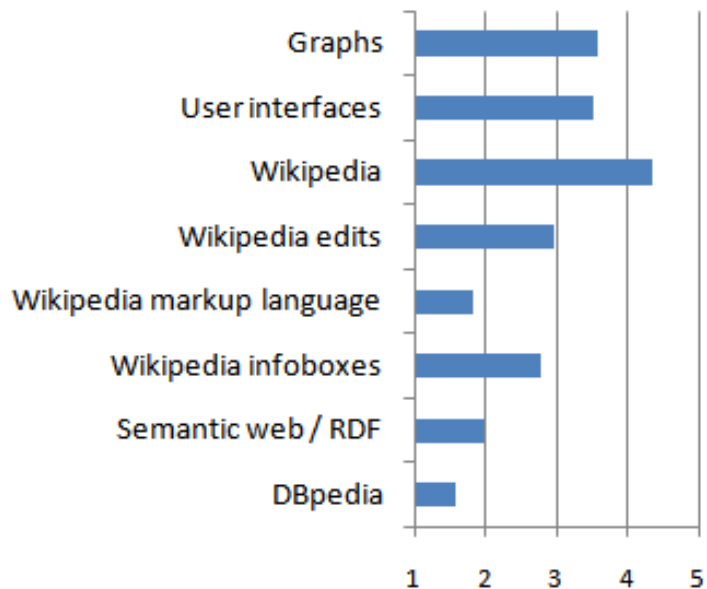
| collegeclubs1 =
| collegecaps1 =
| collegegoals1 =
| years1       =
| clubs1       =
| caps1        =
| goals1       =
| nationalyears1 =
| nationalteam1 =
| nationalcaps1 =
| nationalgoals1 =
| medaltemplates =
| club-update   =
| nationalteam-update =
}}

```

User Study

25 participants

How familiar are you with...



Tasks:

Pick a few related articles

Look at how they are related

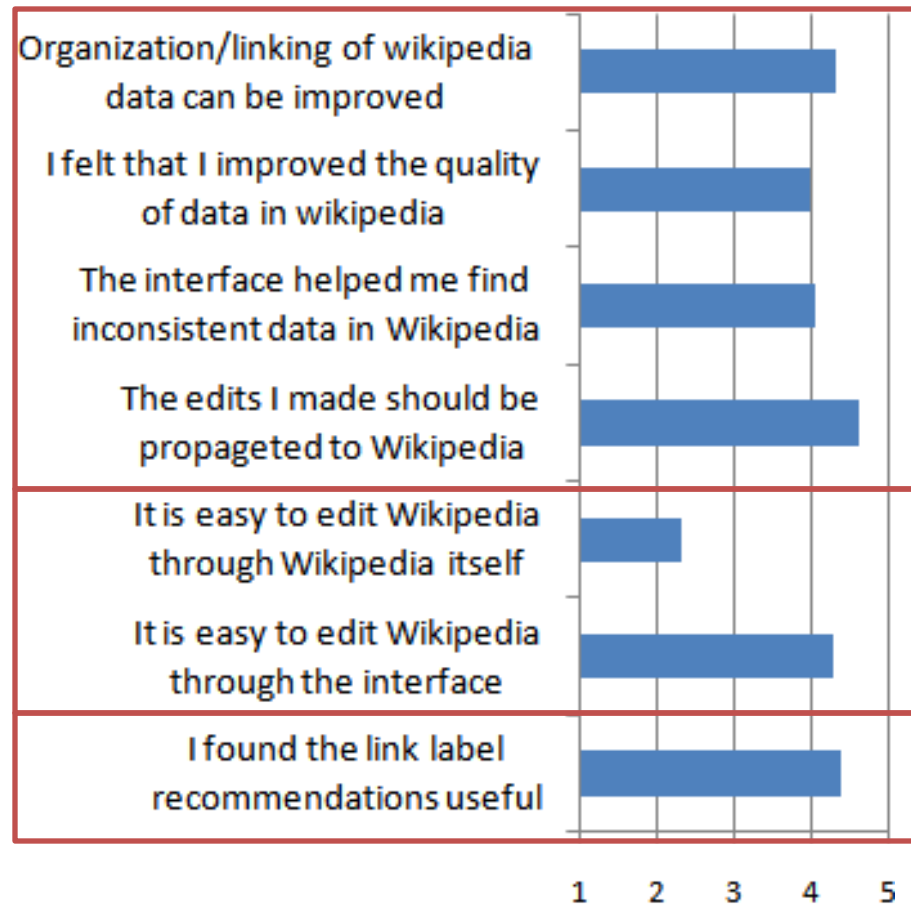
Edit the links as you see fit:

- Through WiGipedia
- Through Wikipedia

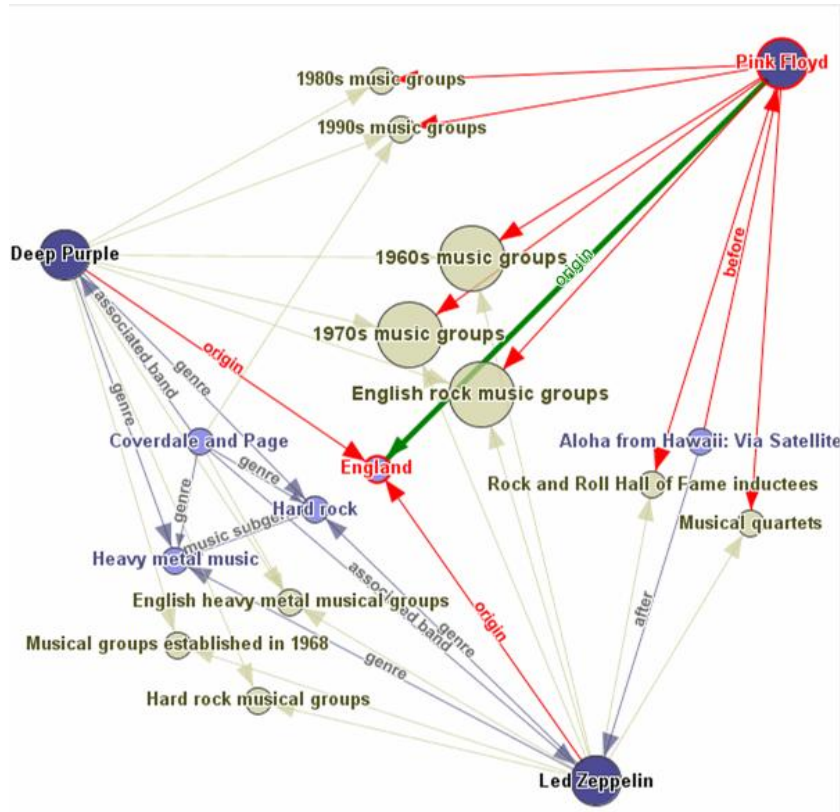
Create a new link and **label** it

Use either **graph or table** or both

User Study



User Study: Graph vs. Table



	● 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

“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)



The image shows a screenshot of a Wikipedia infobox for Citibank, N.A. The infobox contains the following information:

Type	Subsidiary (of Citigroup)
Industry	Finance
Founded	1812
Headquarters	New York City , New York
Key people	Vikram Pandit — CEO ; John Gerspach — CFO ; Douglas Peterson — COO ; William R. Rhodes — Chairman
Products	Financial Services , Retail banking
Website	Citibank.com

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>

Underlying visualization framework

<http://wigis.net>

Contact

alex@cs.ucsb.edu

<http://bostandjiev.com>

