#### Handling Relationships in a Wiki System



Advisor Dr. Chris Pollett Committee Members Dr. Robert Chun Mr. Auston Davis





# Agenda

- Introduction
- Categorization in Wiki Systems
- Problem Statement
- Preliminary Work Summarized
- Relationship Handling System
- Experiments Use Cases and Testing
- Areas of Improvement

Conclusion

Demo



- This project adds a Relationship Handling system to Yioop, an open source content management system to enable relationship specification between the linked wiki pages.
- This allows binary representation of relationships, which does not exist in present day wiki systems.
- This system uses a novel approach by taking into account the linkages between wiki pages and facilitating search functionality on the basis of these links.





- Wiki software enables users to manage content on the web, and create or edit web pages freely.
- Most wiki systems have a feature which allows pages to be grouped together as *categories*.
- This feature assists readers to navigate around related pages to fetch information corresponding to any subject area.
- Categorization is used in many of the enterprise wiki applications, such as Confluence, TWiki and SharePoint, as well as open source-applications like MediaWiki and XWiki.



- Wiki systems support categorization of pages in a very traditional way by specifying whether a wiki page belongs to a category or not.
- The central goal of the category system is to provide navigational links to all Wikipedia pages in a hierarchy of categories which readers, knowing essential—defining—characteristics of a topic, can browse and quickly find sets of pages on topics that are defined by those characteristics.
- Categorization represents unary relationship and is not sufficient to represent n-ary relationships, those involving links between multiple wiki pages.
- To create a category, add [[Category:Parent category name]] to MediaWiki of the wiki page.

# SJSU SAN JOSÉ STATE Categorization in Wiki Systems (Contd.)



Categories are organized as overlapping "trees", formed by creating links between inter-related categories. Partial view of Wikipedia's category system is shown above.



- As we have just seen, Wiki systems have a convoluted structure, which is difficult to manage and retrieve useful information.
- Currently, Yioop specifies linkages between wiki pages but the type of relationship type that binds any two pages is missing.
- Relationship Handling system has been added to Yioop to allow Wiki structure to be explored efficiently, made easily searchable, and navigable on the basis of relationship type between the Wiki pages.
- Use cases for this feature include genealogy, ontology, and dependency identification.



- When categorization was initially added to wiki systems in 2004, there was no mechanism to limit the search results for a particular category. Very large categories caused performance issues.
- Later changes were made to limit the search results to 200 entries per page. Again, for large categories, users had to navigate through multiple pages in order to see all the results. Thus, page by page mechanism also proved impractical.
- In mid-2005, the category table of contents was introduced. With the table of contents, it became possible to navigate through large categories with a few clicks.



- The categorization of Wiki systems has both been analyzed and visualized.
- Holloway et al. [1] compared the top categories and classification structure of Wikipedia 2005 to widely used encyclopedias.
- A more recent study by Kittur et al. [2] analyzed the growth of categories and developed an algorithm to map wiki articles to the top 11 chosen categories.
- Although we see that significant research has been done in the field of wiki categorization structure, the exploration of n-ary relationships is a topic that did not attract much attention among the various studies done on the structure of wiki systems.

# Preliminary Work Summary (Contd.)

- Multiple approaches have been proposed to implement categorization in Wiki systems. These are as follows:
- Intersection
- Manual addition

AN JOSÉ STATE

S

- Pseudo namespace
- Field-value pairs
- IEG proposal using Metrics



# Preliminary Work Summary (Contd.)

- The problems associated with mentioned approaches are as follows:
- Impractical
- Prone to Errors
- Making changes to MediaWiki without including relationship type
- Problem with pages belonging to very large categories
- Inconsistent results, dependent on metrics



- This system works by fetching the type of relationships that link any two wiki pages.
- It stores all relationship types in the database and facilitates transitive relationship exploration and search functionality on the basis of relationship type.
- This feature implements binary relationship between wiki pages, which is sufficient to represent n-ary relationships.
- We combined our approach of adding relationship types linking wiki pages to the approach discussed in the previous slide (Field-value pair), thus, making changes to the wiki markup language syntax.

# SJSU SAN JOSÉ STATE Relationship Handling System (Contd.)

- In Wiki markup language, free links can be created by putting double square brackets around text designating the title of the page that needs to be linked to.
- Thus, [[California]] will be rendered as <u>California</u>. Optionally, a vertical bar () can be used to customize the link title.
- For example, typing [[California|The Golden State]] in the wiki markup will produce a link that is displayed as <u>The Golden State</u> but, in fact, links to California.
- Working on similar approach, relationship type can also form part of media wiki syntax when specifying links between wiki pages as follows: [[capital|Sacramento|Click to go to capital city]] will be rendered as <u>Click to go</u> to capital city and will take the user to Sacramento's wiki page, saving relationship type as capital between California and Sacramento.



- Wiki Parser is responsible for parsing MediaWiki documents, both within Yioop, and when Yioop indexes MediaWiki dumps as from Wikipedia to an HTML equivalent.
- The wiki parser of Yioop fetches each individual MediaWiki internal links present in MediaWiki documents to be able to discover connections between different pages.
- This is done using a regular expression. For example, if page A is related to page B with a link that is displayed as <u>Page B Link</u>, then MediaWiki of document A would have [[PageB|Page B Link]].

 Thus, the presence of these types of internal links can be extracted using a regular expression. Once we have all the links between the pages, it can be saved to the back end, and later used for information retrieval of linkages between different pages.



• Consider a test group, named 'NewTest' with three Wiki pages: Main, test1new, and test2new. The links between them are summarized as below:

Relationship Type	From Page	To Page
parent	Main	test1new
parent	Main	test2new
parent	test1new	test2new
child	test1new	Main
child	test2new	test1new



# "What Links Here" feature (Contd.)

# COOS - Admin [Feeds and Wikis]

Account Access	
Manage Account	test1new Page at NewTest Group 🔹
Manage Users	test1new Page at NewTest Group Places
Manage Roles	test1new Page at NewTest Group
Crawls	Edit test1new Page
Manage Crawl	What links to tost1 now Page
Manage Classifiers	What initial to test inew Page
Page Options	test1pew Page Discussion
Results Editor	NewTest Group Page List
Search Sources	NewTest Group Feed
Web Scrapers	Recent Pages
SUCSIS SU	SOLSES OF STATES SOLSES OF STATES OF
le se le s	zuele zu



# "What Links Here" feature (Contd.)





- To delve further into how wiki pages link with each other, i.e., information about their relationship types, "What Relates To" feature is implemented.
- It works in conjunction with the wiki parser to extract the relationship type by creating a link extractor regular expression which takes into consideration the relationship type.
- For example, if page A is related to page B with a link that is displayed as "Page B Link," and a relationship "Parent," then MediaWiki of document A would have [[Parent|PageB|Page B Link]].
- The fetched values, i.e., relationship type and the linked page are then stored in the database and can be displayed to check what type of relationships link these page to other pages.



# "What Relates Here" feature (Contd.)



ccount Access	
age Account	What relates to test1new Page
ge Users	test1new Page at NewTest Group Places
je Roles	test1new Page at NewTest Group
vle	Edit test1new Page
de Crawl	test1new Page History
te Classifiers	What links to test1new Page
Ontions	What relates to test1new Page
	test1new Page Discussion
ts Editor	NewTest Group Page List
n Sources	NewTest Group Feed
Scrapers	Recent Pages
	וצחיע בוצחיע בוצחיע בוצחיע בוצחיע בוצחיע בוצחיע בוצחיע בוצח



# "What Relates Here" feature (Contd.)





# "What Relates Here" feature (Contd.)



#### Crawls



- The Group Page list in Yioop displays all the wiki pages present in a particular group. The advanced search feature has been added to it by displaying a search text box, which enables search on the basis of relationship type chosen and any particular page.
- The user can choose multiple relationship types and the results would display all wiki pages linked to that particular page with the chosen relationship type.
- These relationships can be further explored to check which pages link "from" and which pages link "to" the wiki page selected by the user.

Since Since



Account Access	- Ac	min [Feeds and Wikis]	
Manage Account		New lest Gloup Fage List	
Manage Users		NewTest Group Page List	
manage Roles		Search group page titles Go Advanced	ł
Crawls			-
Manage Crawl			
Manage Classifiers		Main Testnew Main pagel	
Page Options		loonow main page:	
Results Editor			
Search Sources		test1new	
Web Scrapers		This is test1 new page!	8515
			s.vin
Social			28
Manage Groups		test2new This is test2 new pagel	กรรร
Feeds and Wikis		This is took non page.	5 20
Mix Crawls	SC	NE NE NE NE NE NE NE NE	NC NC
ភទ]ຮບລອງຮບລອງຮບ	JS SJ	suasisuasisuasisuasisuasisuasisuasisuas	ເອຊີນ ເອ
ខិតថាខុខ្មែតិបទខ្មែតិបទ	SIS S	ມຂໄຂ້ເຈັບຂໄຂ້ເຈັບຂໄຂ້ເຈັບຂໄຂ້ເຈັບຂໄຂ້ເຈັບຂໄຂ້ເຈັບຂ	ເຂັສັບຂຸເຂັສັບ
ns NS NS NS	SC	nse	SU SU DS
๛รารบ๛รารบ๛รารเ	JS SI	SUW SISAM SISAM SISAM SISAM SISAM SISAM SISA	เจรารบจราร



Account Access	NewTest Group Page List •
Manage Users	New Test Crown Dere List
Manage Roles	New lest Group Page List
Crawls	Search group page titles Advanced
Manage Crawl	
Manage Classifiers	parent
Page Options	Sibling Enter group page title Go
Results Editor	sister -
Search Sources	•
	rslē సිuzlē సිuzlē సිuzlē సිuzlē సිuzlē సිuzlē సొuzlē సొuzlē స



	<b>\$000</b> - A	dmin [Feeds and Wikis]
	Account Access	NewTest Group Page List
	Manage Users	
	Manage Roles	NewTest Group Page List
		Search group page titles Advanced
	Crawls	
	Manage Crawl	mothor
	Manage Classifiers	parent
	Page Options	sibling Main Go
	Results Editor	sister
	Search Sources	
	Web Scrapers	
	Social	test1new
	Manage Groups	
	Feeds and Wikis	test2newSUCT SISU
	Mix Crawls	ns(sୁଦ୍ଧର)
US		Sisting stands and stands an







### Basic flow of processes in RH system





- The experiments conducted on the Relationship Handling (RH) system give an idea about the working of the system under ideal conditions. Here, an assumption is made that the content added to wiki pages adheres to MediaWiki syntax.
- These experiments are divided into four parts:
  - Experiments conducted to test the working of all new features added to RH system – Part of Demo
  - o A/B Testing
  - o Performance Testing
  - Experiments conducted on use-case scenarios of the RH system



- This testing, also known as split testing, involves comparing two different versions of a web page to see which one performs better.
- The two variants are shown to similar visitors and their responses are recorded. The one that provides a better conversion rate, wins.

Account Access Manage Account Manage Users Manage Roles Crawls	NewTest Group Page List           NewTest Group Page List           Search group page titles         Advanced	Account Access Manage Account Manage Users Manage Roles Crawls	NewTest Group Page List           NewTest Group Page List           Search group page titles         Advanced	
Manage Crawl Manage Classifiers Page Options Results Editor	brother child husband	Manage Crawl Manage Classifiers Page Options Results Editor	brother child husband - Enter from group page title Go	
isus sisus s	insiz Susiz Sus		Susla Susl	
Sonsis So	S S C S S C S S C S C S C S C S C S C S	SUSSICE SUSSIC	Sisna sisn	



## Experiments – A/B Testing (Contd.)

The results of A/B testing are as follows:





• This testing was done by adding multiple relationship links to a wiki page and then creating that page. The time of creation was noted for a wiki page by varying the number of links.



Performance Testing



- The developed system has been found really useful in the field of genealogy to organize the family tree structure better and understand it efficiently.
- This system has been tested on Dr. Pollett's family tree containing around 300 members.
- Screenshots depicting the usefulness of this system to manage large family trees easily are as follows:







The Group Page List under Genealogy group displays all members of Dr. Pollett's family tree.

Genealogy Group Pag	ne List			
Search group page titles		Go		
Absalom_Pollett [Edit] Absalom Pollett Birth: 1876-09-18				
Adrienne Watson [Edit] Adrienne Watson Birth: 0000-00-00				
Agenard J. Belcourt [Edit] Agenard J. Belcourt Birth: 1864-00-0			71e t 195	
Albert_Cranford [Edit]			2	JSIS SUSI



To test Dr. Pollett's relationship to other members of the family, search for "Christopher Pollett" in the search box.

	- Genealogy Group Page List	
	Genealogy Group Page List	
	Christopher Go	
	Create Page: Christopher	
	Christopher_John_Pollett [Edit] Christopher John Pollett Birth:	
SISU SSISU SSISU SSISU	is such such a such such a such such a such	Signal Signa
SUSIS		SILISIS SILISIS



Dr. Pollett's Wiki page is as follows:



Sincle Si



Dr. Pollett is related to Mr. Harry Fraser Lancaster Pollett with the relationship son. His father's wiki page can be explored further by clicking on the link displaying his name.



SISUSIS SISUS SISUSIS SISUSIS



To check all people with whom Dr. Pollett is linked, choose "What links to Christopher John Pollett Page" from the drop down.





To check all relationship types that link Dr. Pollett are as follows.





Any relationship can be further explored to check "from" and "to" relations.



- The developed system can also be used to identify entities that have no dependencies or those entities that are not dependent on any other entities.
- This can be done using a crawler to figure out pages with no links "to" and "from" other pages.
- For example, a project with multiple components where execution of one component is dependent on another. We can find relationships between various components and figure out which needs to be executed first.



- The areas of improvement includes scenarios where two-way relationship exists between two entities. For example, if A is father to B, then B is son/daughter to A.
- In this system, both relationships have to be specified separately. Improvements can be made where if one side of such relationships is specified, other sides get identified automatically.
- In addition to this, future scope includes creating category and sub category pages using the syntax: [[category name| |Text to be displayed]] and creating multiple relationships to a wiki page simultaneously by using check boxes.



# Conclusion

- In this project, a new module called Relationship Handling system is developed for Yioop.
- This module is responsible for storing the information in a manner such that it can be later used for understanding the links between wiki pages and searching based on relationship types.
- This system employs an approach of storing the relationship types and links between wiki pages in the database to be used for exploring relationships and facilitating search features.

• It is very useful to understand the convoluted structure of wiki systems and letting the user explore content easily and effectively.



# References

[1] T. Holloway, M. Bozicevic, and K. Borner, "Analyzing and visualizing the semantic coverage of Wikipedia and its authors," *Complexity* 12, 2007, pp. 30–40.

[2] A. Kittur, E. H. Chi, and B. Suh, "What's in Wikipedia? mapping topics and conflict using socially annotated category structure", in *Proc. 27th Annual CHI Conference on Human Factors in Computing Systems (CHI'2009)*, New York, USA, 2009, pp. 1509–1512.

[3] arXiv.org, "Evolution of Wikipedia's category structure," 2012. [Online]. Available: https://arxiv.org/abs/1203.0788. [Accessed: 20- Nov- 2016].

[4] Ontology. (2016). Retrieved from https://en.wikipedia.org/wiki/Ontology\_(information\_science).

[5] L. Muchnik, R. Itzhack, S. Solomon, and Y. Louzoun, "Self-emergence of knowledge trees: Extraction of the Wikipedia hierarchies," Physical Review E 76, 2007.

[6] MediaWiki links. (2015). Retrieved from https://www.mediawiki.org/wiki/Help:Links#Internal\_links.



# References (Contd.)

[7] M. Bergman, "Wikimedia architecture" (PDF). Wikimedia Foundation. [Accessed 27-Sep-2016].

[8] "Announcement of Wiktionary's creation," meta.wikimedia.org. [Accessed 10-10-2016].

[9] arXiv.org, "Connecting every bit of knowledge: The structure of Wikipedia's first link network," 2012. [Online]. Available: https://arxiv.org/abs/1605.00309. [Accessed: 05- Nov- 2016].

[10] Wiki markup. (2015). Retrieved from https://en.wikipedia.org/wiki/Help:Wiki\_markup.

[11] Seekquarry.com, "Resources," 2015. [Online]. Available: https://www.seekquarry.com/p/Resources. [Accessed: 09- Sep- 2015].

#### Demo

#### SAN JOSÉ STATE UNIVERSITY *powering* SILICON VALLEY

