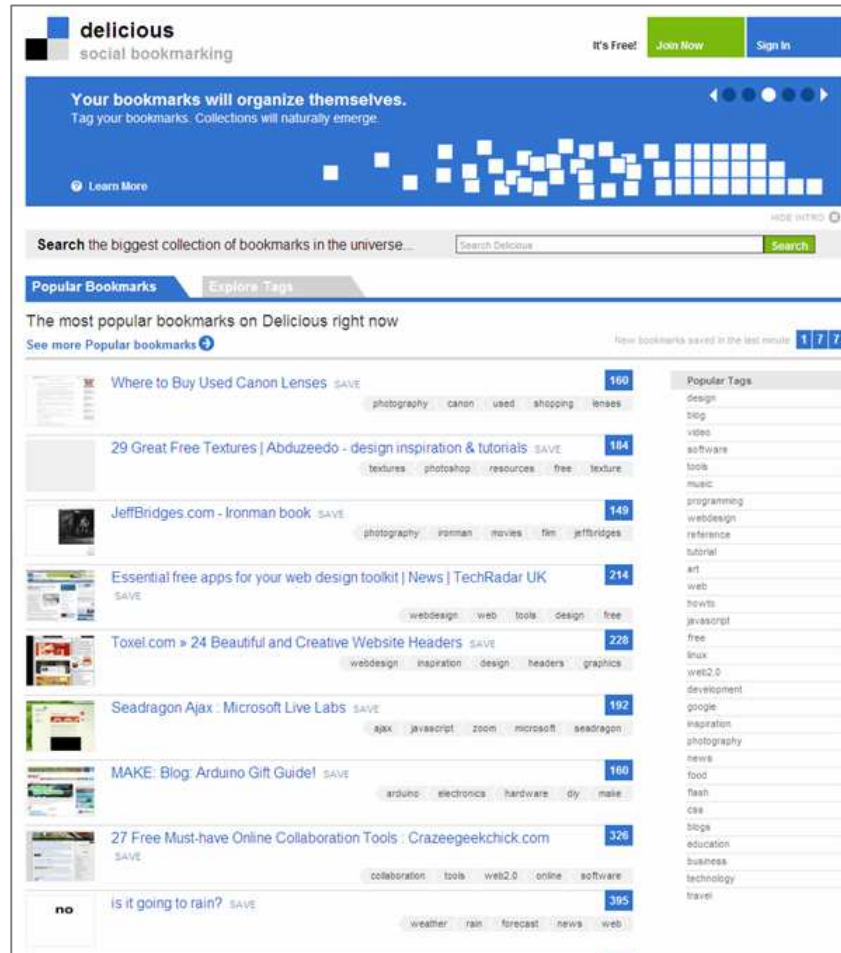


# Collective User Behaviour and Tag Contextualisation in Folksonomies

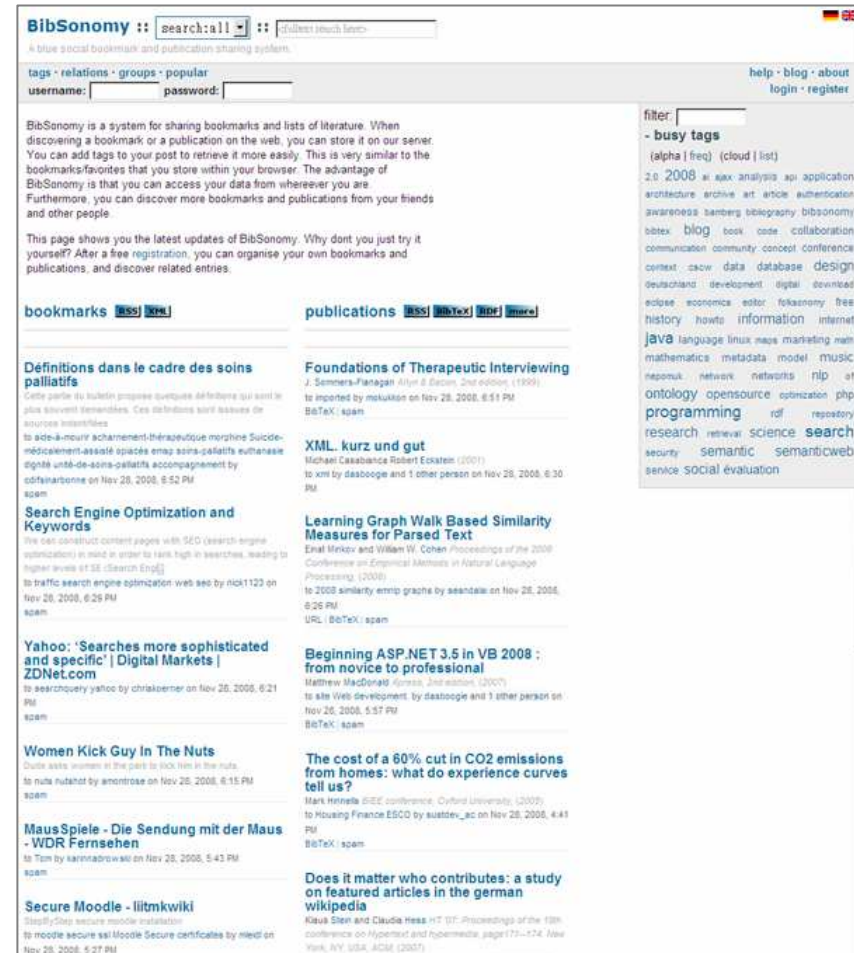
by Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt

Intelligence, Agents, Multimedia Group  
School of Electronics and Computer Science  
University of Southampton

# Introduction

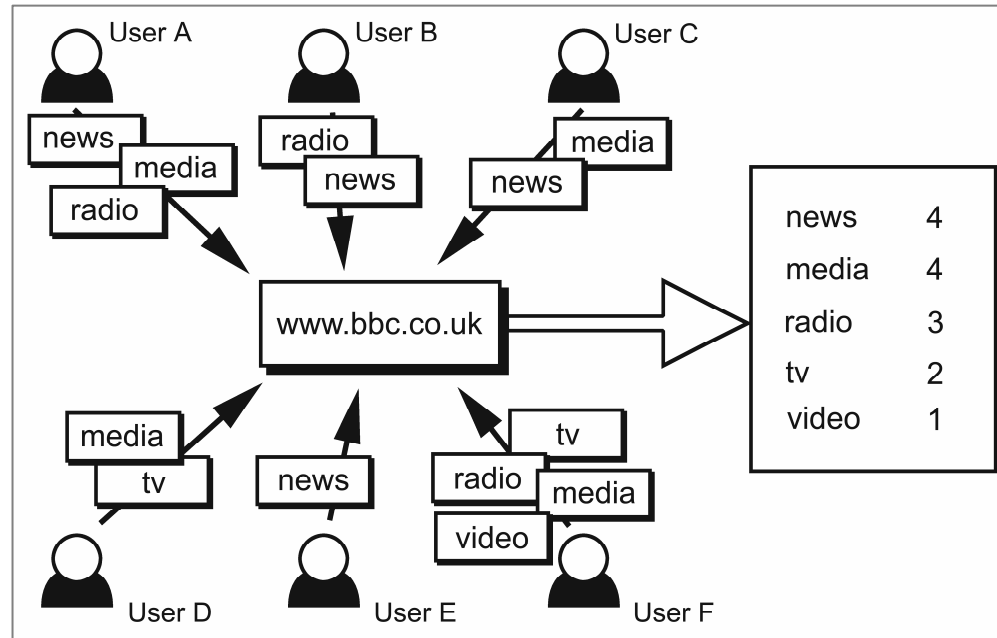


Delicious



BibSonomy

# Introduction



An example of  
collaborative tagging

A Tag Cloud in Delicious

gging blogs books business china community co  
development diy download drupal education  
firefox flash flickr food forum free freeware fo  
hardware health history home howto html humor illu  
internet iphone japan java javascript jobs learning li  
math media microsoft mobile mp3 music news  
raphy photos photoshop php plugin politics portfo  
ils recipe recipes reference research res  
/ seo sga shop shopping slash social software  
bread travel tutorial tutorials tv typography ubun

## The Problem of Ambiguity

- ◆ A tag can be used to refer to different concepts
- ◆ Understanding the meaning of a tag is context-dependent
- ◆ Limits the effectiveness of collaborative tagging and causes problems when retrieving documents
- ◆ Example:  
*sf* → *San Francisco*, or  
*sf* → *Science Fiction*?

# Background and Motivations

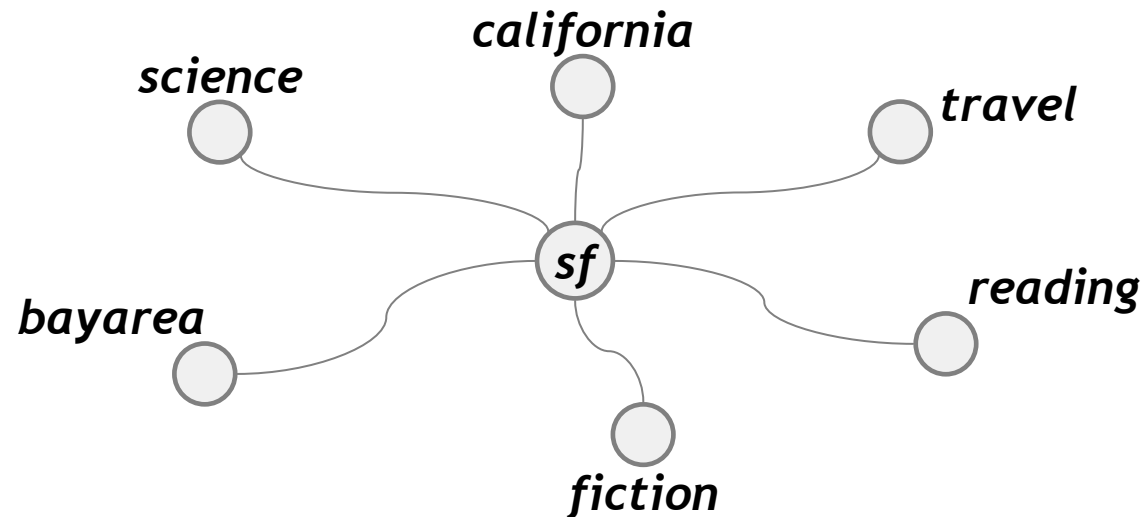
The screenshot shows the Delicious website interface. At the top, there's a navigation bar with 'delicious' logo, 'Home', 'Bookmarks', 'People', and 'Tags' tabs. A search bar is on the right. Below the navigation bar, the main content area is titled 'Popular sf Bookmarks' with a sub-header 'Recent | Popular'. A link says 'See all sf bookmarks or go to your Bookmarks.' Below this, a list of bookmarks is displayed, each with a title, a 'SAVE' button, a count in a blue box, and a list of tags. The bookmarks are:

- Futurere (JBR Precoglang)** (306) Tags: literature, futurism, scifi, language, sf. First saved by: snfg.
- 現代SFお奨め作品一覧。 - Something Orange** (48) Tags: book, sf. First saved by: yoyogi.
- Heinlein's Fan Mail Solution** (126) Tags: writing, writers, heinlein, formletters, sf. First saved by: magnetbox.
- Ten Truly Underrated Sci-Fi Movies - VideoHound's Golden Movie Retriever** (329) Tags: SF. First saved by: nnschiller.
- Plan59 :: 1950s Space Art** (156) Tags: images, posters, retro, vintage, sf. First saved by: billmspg.
- Death Star over San Francisco // Current** (414) Tags: sci-fi, video, illusion, movies, sf. First saved by: nedward.
- 36 Hours in San Francisco's Mission District - NYTimes.com** (29) Tags: travel, sf. First saved by: tshelly.
- SF初心者これだけは読んどけ! : アルカン速報** (18) Tags: まとめ, book, sf. First saved by: deyas.

On the right side, there's a 'Tags' section with 'Related Tags' (11) including: studio, français, mission, bayarea, blog, usa, toread, \*stuff, tv\_show, sidebar, gene\_wolfe. At the bottom, there's a footer with links: delicious, about, blog, terms of service, privacy policy, copyright policy, forums, support, and a 'What's new?' button.

## Contextualisation of Tags

- ◆ Putting tags in context by studying their relations with other tags, users and documents
- ◆ Limitation of large scale clustering of tags:

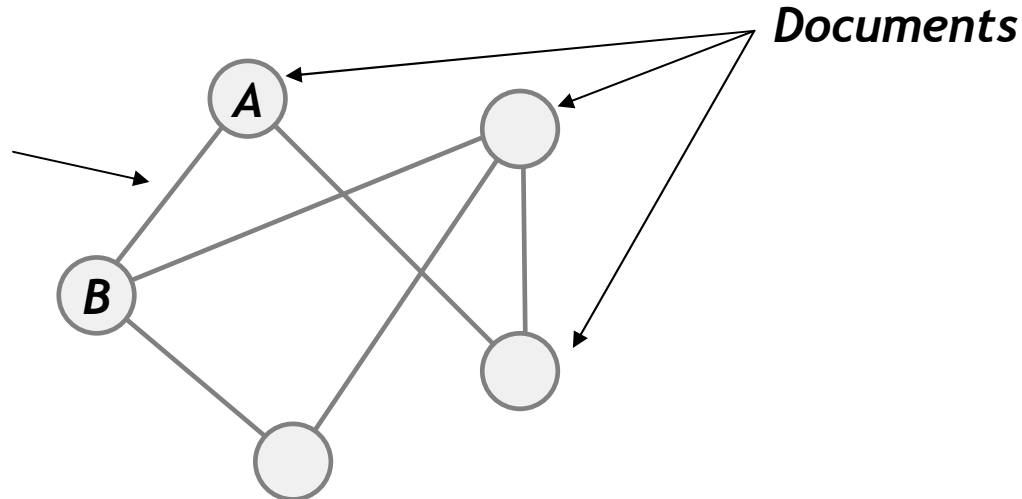


## Contextualisation of Tags

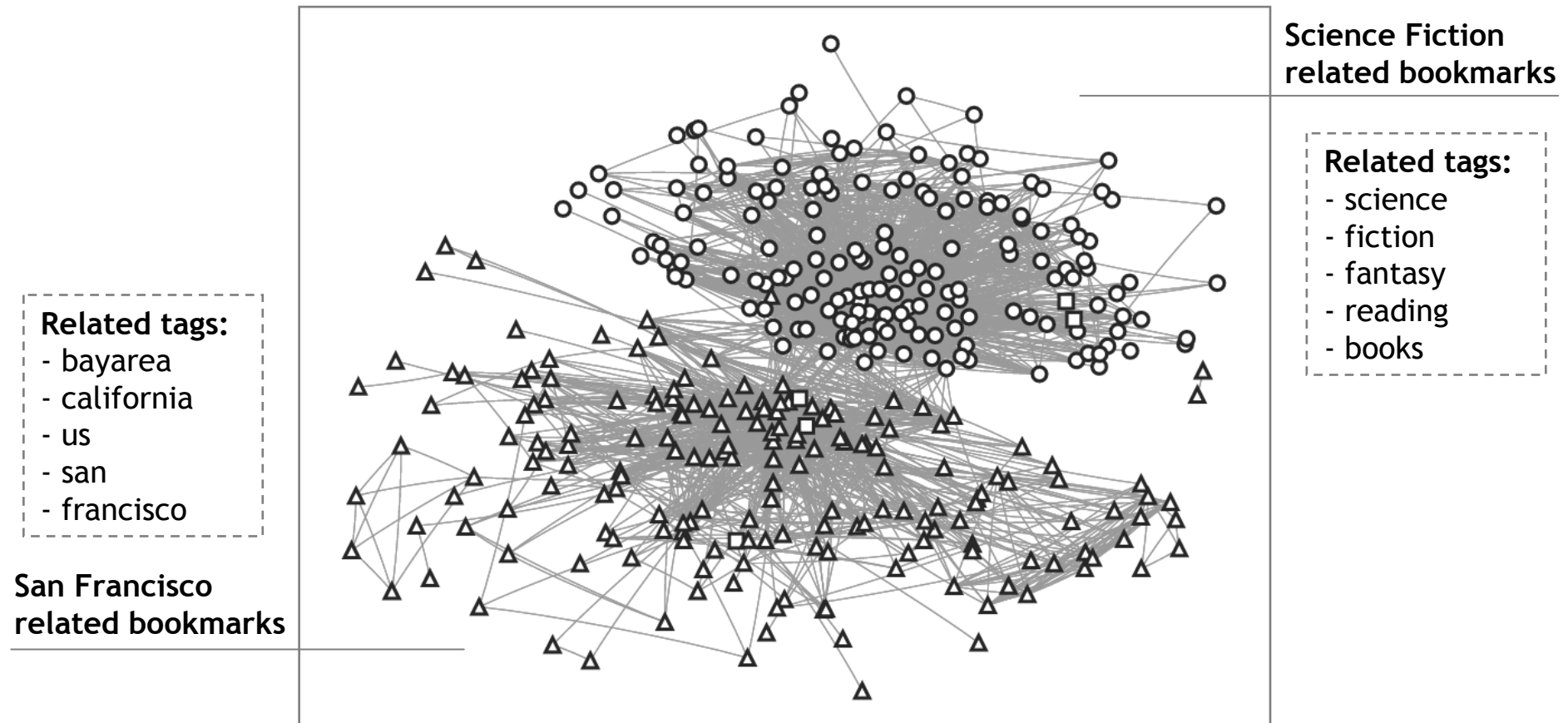
- ◆ Instead, we look at a particular tag, construct networks based on the users' behaviour

An edge exists if some users assigned the tag to both A and B.

(Compare to the case in which they share some tags)



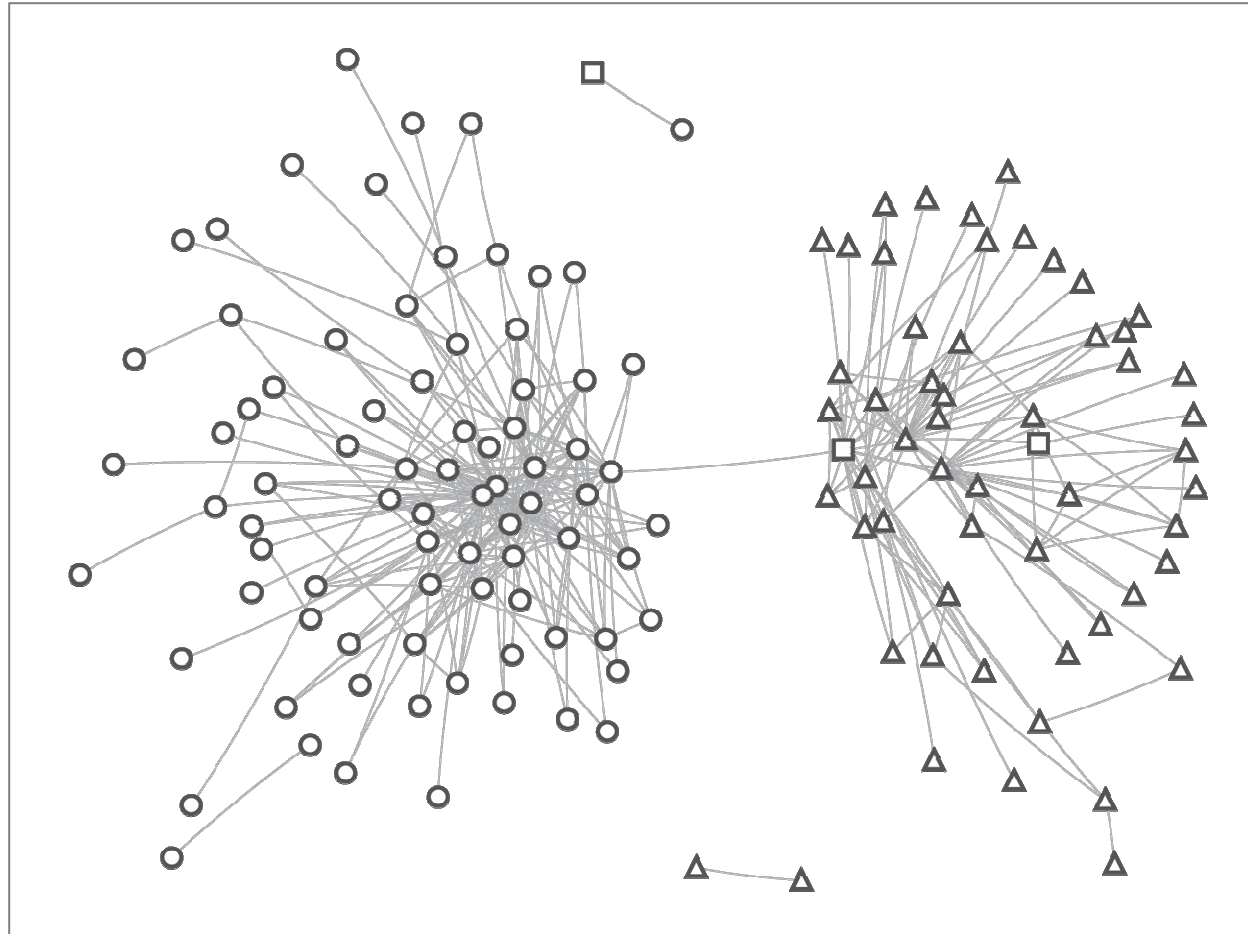
# Preliminary Results



Network of bookmarks with edges weighted by number of users using the tag *sf* on the bookmarks.  
(Data obtained from Delicious)



## Preliminary Results



The same network with edges weighted less than 2 removed.

## Summary

- ◆ Documents can be clustered by considering only the pattern of how the users use a particular tag
- ◆ Users tend to be consistent when using a tag (they tend not use it ambiguously)
- ◆ Modelling folksonomy data in this way gives insight into how the tags are actually used

# Preliminary Results

## More Examples

Tag	Context	Tags Extracted
<b>tube</b>	YouTube	tube, youtube, video, funny, videos, fun, cool, music, feel.good
	Vacuum Tubes	tube, audio, electronics, diy, amplifier, amp, tubes, music, elect
	Underground	tube, london, underground, travel, transport, maps, map, uk
<b>bridge</b>	Design Pattern	bridge, programming, development, library, code, ruby, tools
	Card Game	bridge, games, cards, game, imported, howto, conventions
	Networking	bridge, networking, linux, network, howto, software, sysadmin
	Architecture	bridge, bridges, structures, engineering, science, physics
<b>wine</b>	Linux Software	wine, linux, ubuntu, howto, windows, software, tutorial, emulation
	Beverage	wine, food, shopping, drink, reference, vino, cooking, alcohol, blog

## Research Questions

1. Does the social context provide better information for understanding the semantics of tags?
2. Which method of modelling the data in a folksonomy best represent the associations between documents/users/tags?
3. How can we represent the different contexts extracted?

## Different Kinds of Networks

- ◆ User-based tag networks / document networks  
(Social context considered)
- ◆ Tag-based user networks / document networks  
(Similar to the vector model in IR)
- ◆ Tag similarity networks  
(e.g. Calculate similarity of the contexts of two tags)

## Data Preparation

- ◆ We collected data of 50 tags (associated bookmarks, users and tags) from Delicious
- ◆ Recruited users to manually classify bookmarks into different contexts
- ◆ Compare the output of clustering process done on different networks
- ◆ Performance measures: precision, recall, coverage, etc.

## Other Possible Evaluation Methods

- ◆ Comparing with document clustering techniques e.g. Clusty



The screenshot shows the Clusty search engine interface. At the top, there's a navigation bar with links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. A search bar contains the text 'sf', and there are buttons for 'Search' and 'advanced preferences'. Below the navigation bar, there's a sidebar with tabs for 'clusters', 'sources', and 'sites'. Under 'clusters', there's a list of results: 'All Results (254)', 'Fiction, Science (29)', 'Bay, San Francisco (31)', 'Photos (14)', 'Meetings, San Francisco (11)', and 'Blog (11)'. The main content area shows 'Top 253 results of at least 43,377,536 retrieved for the query sf (details)'. Below this, there's a section for 'Current stock trading for Stifel Financial Corp. (SF)' with a table showing stock data. The table has columns: SYMBOL, LAST, CHANGE, OPEN, and PREV CLOSE. The data row shows: SF, 43.01, +0.13 (0.30%), 42.50, 42.88. Below the table, there's a list of search results, with the first one being '1. The SF Site: The Best in Science Fiction and Fantasy'.

SYMBOL	LAST	CHANGE	OPEN	PREV CLOSE
SF	43.01	+0.13 (0.30%)	42.50	42.88

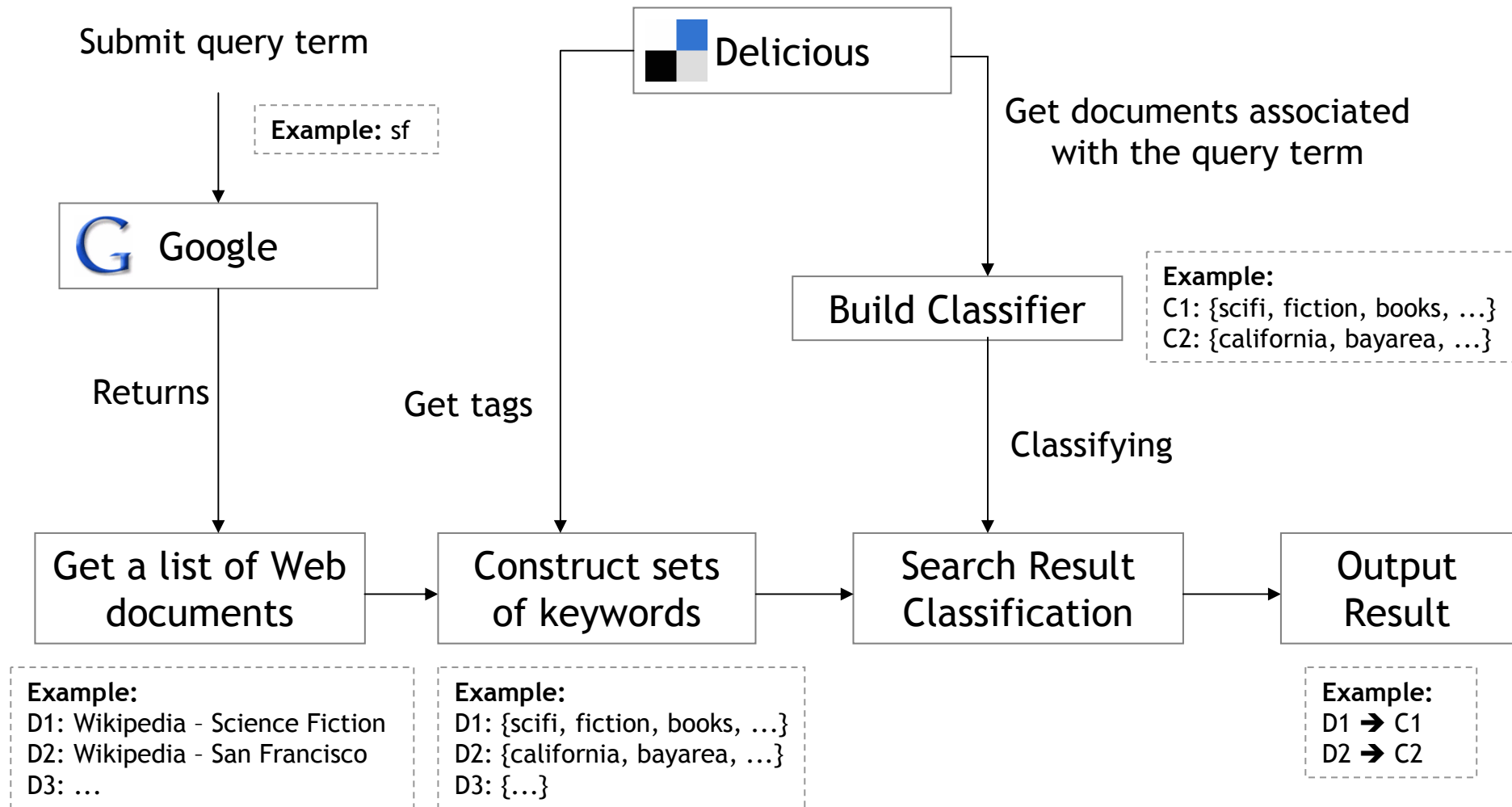
- ◆ Comparing the results with WordNet or other thesaurus

## Web Search Result Classification

- ◆ We applied the result on classifying Web search results returned by search engines
- ◆ New or domain-specific meanings of a keyword can be identified (e.g. tube, bridge)
- ◆ Paper: *A k-Nearest-Neighbour Method for Classifying Web Search Results with Data in Folksonomies*  
(Session 10-A-WI-4, 10 Dec, 10:45)



# Applications



# Conclusions

- ◆ Outlined our research on tag contextualisation based on collective user behaviour
- ◆ Social contexts should be considered when trying to understand how tags are used
- ◆ Results can be applied to applications such as enhancing Web search or recommendation systems

**Thank You!**

Albert Au Yeung  
cmay06r@ecs.soton.ac.uk  
<http://users.ecs.soton.ac.uk/cmay06r/>