

# Соціальні закладки, “народна класифікація” – складні мережі

## Social Bookmarks, **Folksonomies** – Complex Networks

Короткий огляд, історія виникнення,  
математичні основи, тенденції розвитку

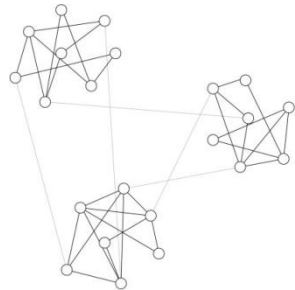


Тернопільський технічний  
державний університет  
імені Івана Пулюя,  
Кафедра комп'ютерних наук

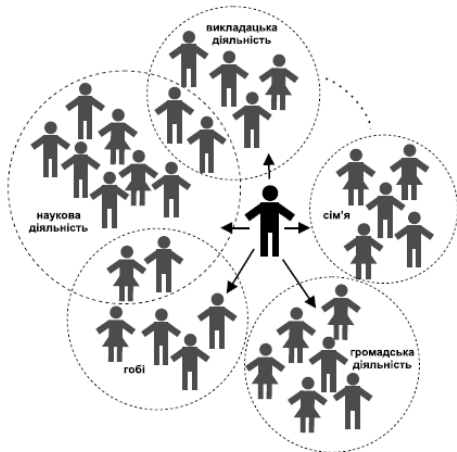
# Приклади систем (явищ), що мають форму складних мереж (Complex Networks)

- Інтернет, www.
- Нейронні мережі.
- Мережі метаболізму, харчування.
- Дистрибуція товарів та послуг (MLM).
- Транспортні мережі, постачання (газу, нафти), розподільчі (пошта).
- Соціальні мережі (odnoklasniki.ru, vkontakte.ru).
- Мережі співпраць (profeo.com.ua, linkedin.com).
- **Соціальні закладки (del.icio.us, flickr.com, furl.net).**
- Мережі цитування наукових статей (scholar.google.com, portal.acm.org, sciencedirect.com).

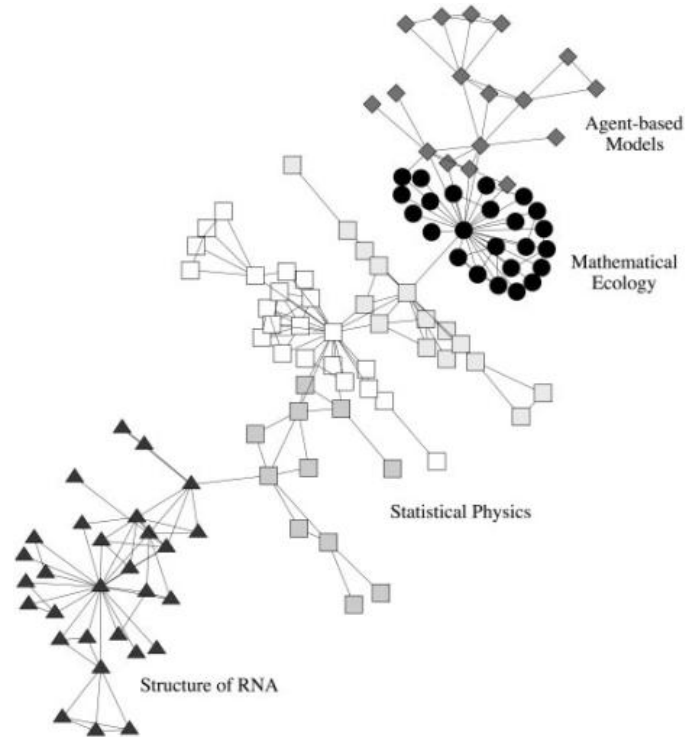
# Математичні основи складних мереж – теорія графів



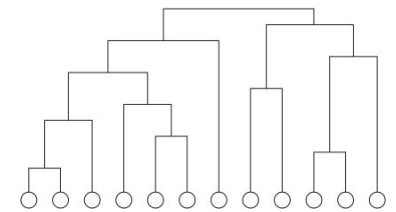
Схематичне представлення мережі, що утворює ком'юніті



Соціальна діяльність людини

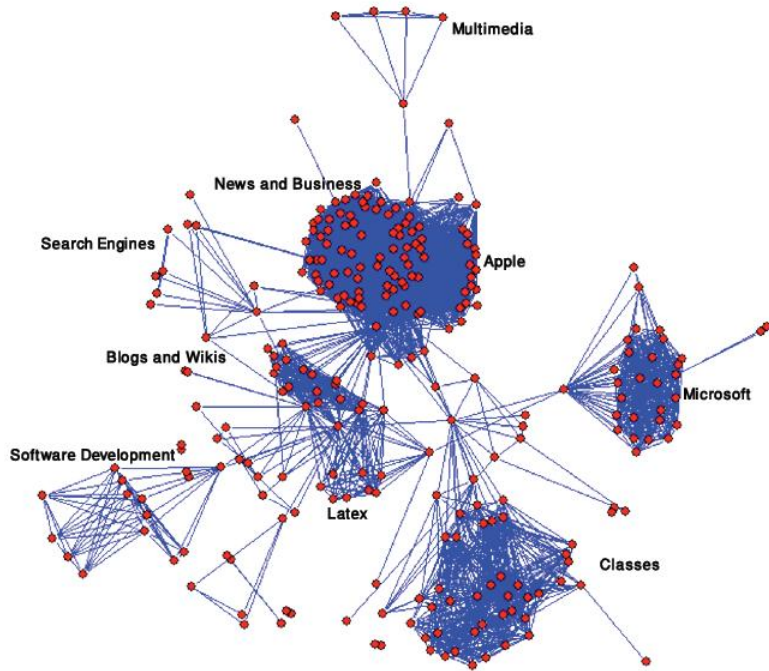


Мережа співпраці між різними науковими напрямками та школами

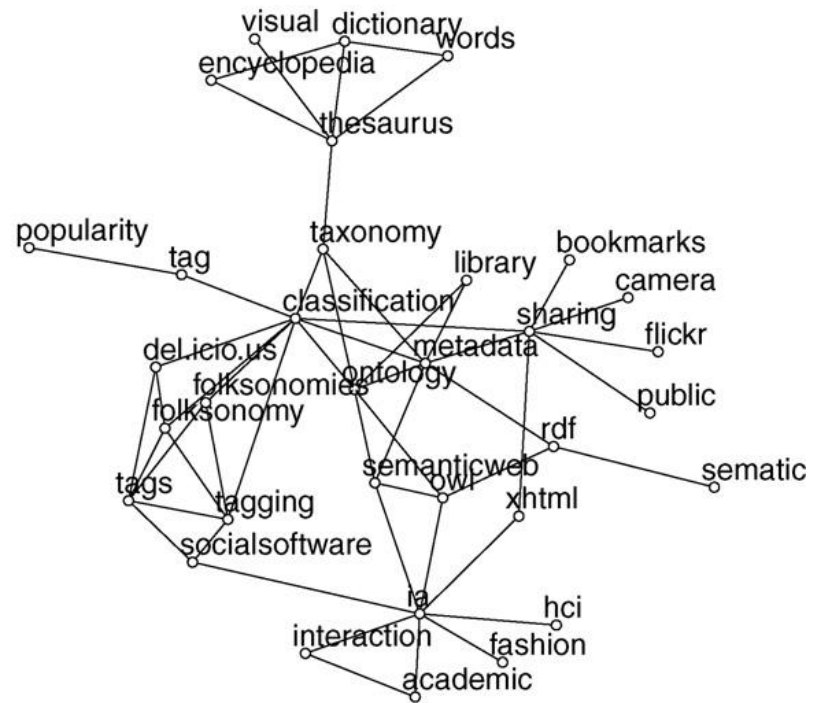


Приклад малого ієрархічного кластерного дерева

# Математичні основи складних мереж – теорія графів

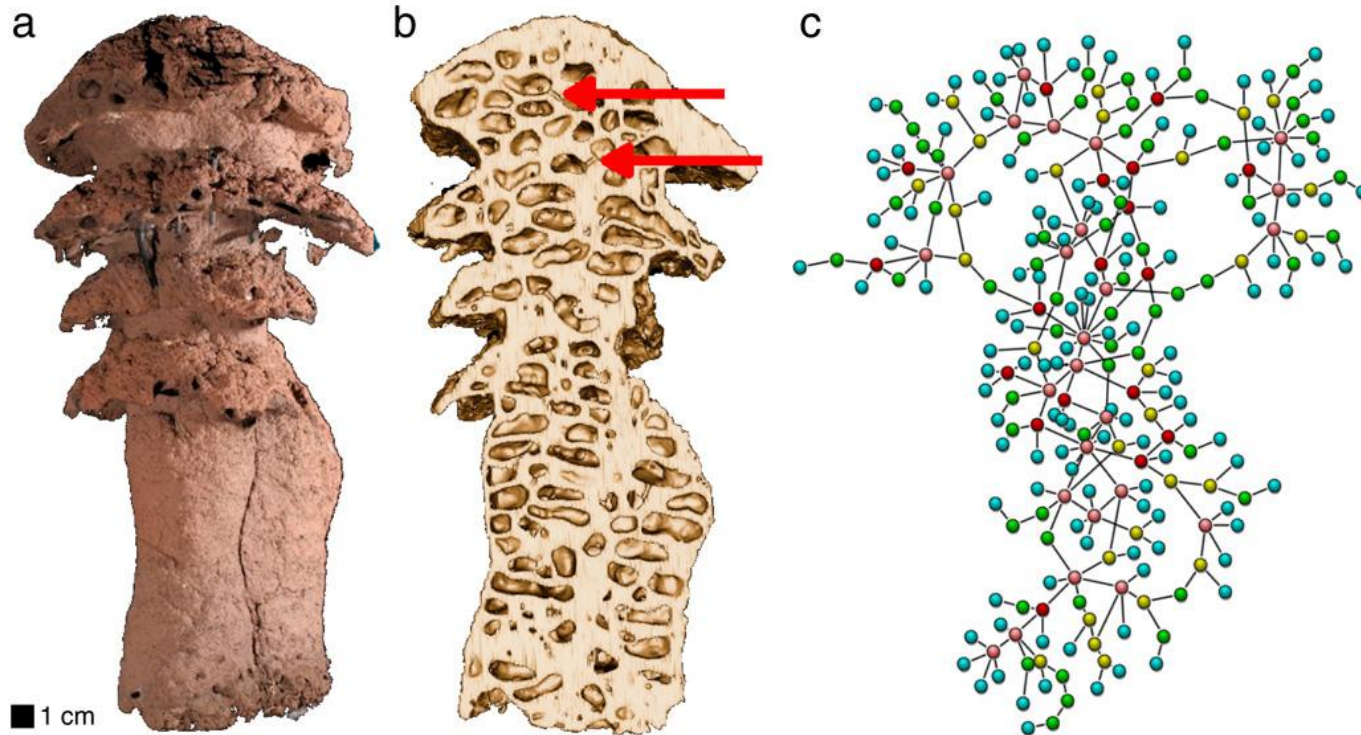


Графічна топологія, яка згенерована програмою Ражек



Графічне представлення з сайту del.icio.ua, яке користувачі асоціюють із терміном “ontology”

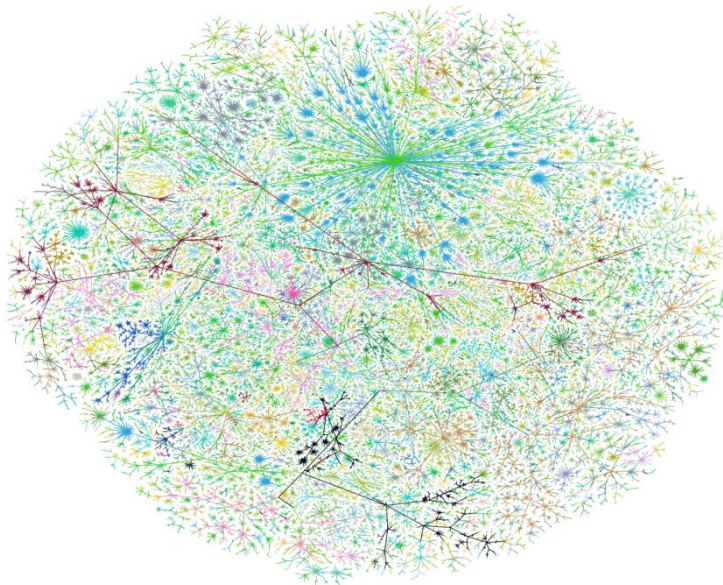
# Математичні основи складних мереж – теорія графів



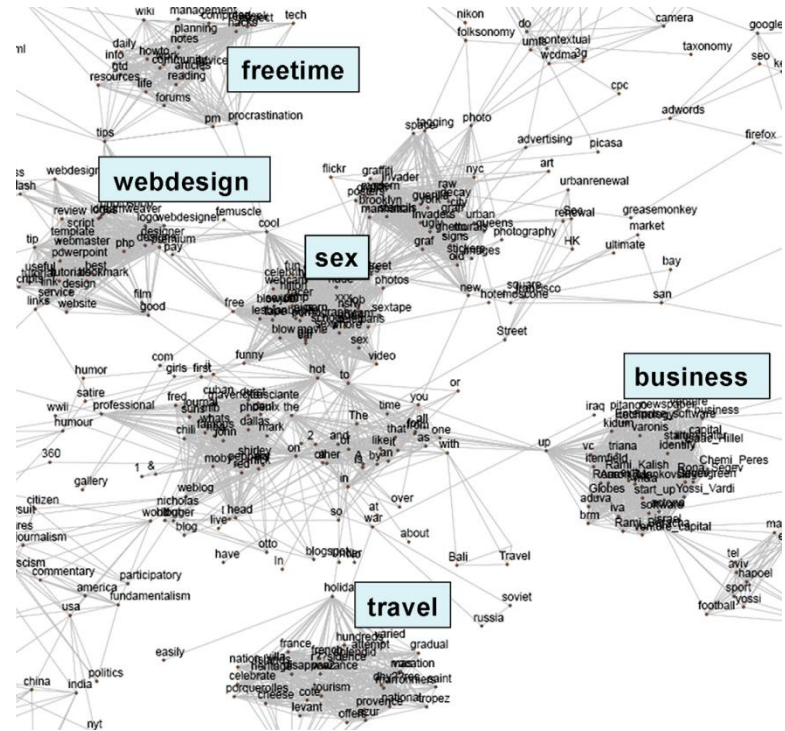
**Зліва:** зображення термітного гнізда; **посередині:** томографічний розріз цього ж гнізда; **справа:** репрезентація комірок та галерей як мережі (графа), де вершина - це комірки, а ребра графа - коридори сполучення. Колір вершин відображає ступінь сполучення комірок.

*A. Perna et al. / Physica A 387 (2008) 6235–6244*

# Математичні основи складних мереж – теорія графів



Візуальне  
представлення мережі  
інтернету

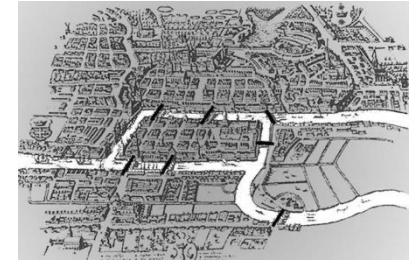


Del.icio.ua, приклад  
відображення тегів -  
наглядна демонстрація  
появи кластерів

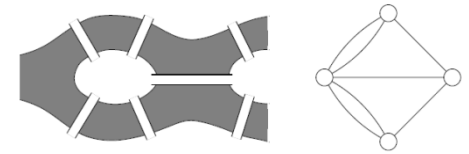
# Коротка історія виникнення теорії графів



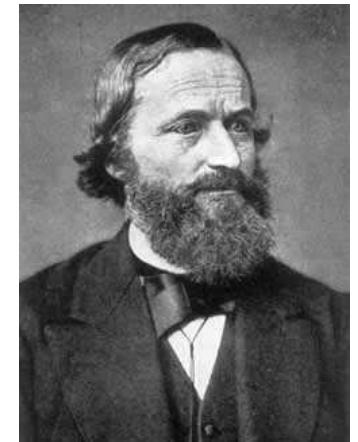
- Ойлер (Leonard Euler, 1707-1783) у 1736 р. сформулював задачу про сім мостів Кьонігсберга (найстаріший граф).



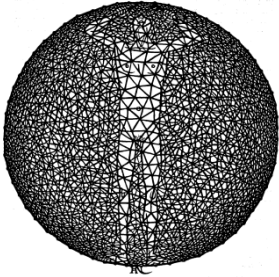
- Френсіс Гасрі (Francis Guthrie, 1831-1899) у 1852 р., розфарбовуючи карт, графств Англії, сформулював задачу про чотири кольори.



- Густав Роберт Кірхгофф (Gustav Kirchhoff, 1824-1887) у 1845 р. опублікував правила для електричного кола.



# В ХХ столітті застосування теорії графів поширилося на інші галузі науки



- Соціологія встановлює зв'язки між членами суспільства. Представлення таких зв'язків у вигляді графів дозволяє вивчити їх кількісно.
- С. Мілгрем 1967 р. – “шість ступенів розділення” попередник ефекту “тісного світу”, що характерне для більшості складних мереж.
- Пауль Ердош (1913-1996) – число Ердоша як приклад мережі співпраці.
- Альфред Рені (1921-1970) – разом з П.Ердошем ввели поняття класичного випадкового графа або випадкового графа Ердоша-Рені
- Згодом теорію графів використовували в інформатиці, кібернетиці, біології та фізиці



# Математичні моделі мереж – три головних види графів

Складні мережі. Головач Ю., Олемской О., К. фон Фербер та інші // Журнал технічних досліджень т.10, № 4 (2006). 247-289

Однією з головних характеристик мережі є розподіл ступенів вузлів  $P(k)$ , що визначається як імовірність того, що вузол  $i$  має ступінь  $k_i = k$ . Під час дослідження виявилось, що мережі, які характеризуються різними  $P(k)$ , демонструють дуже різноманітну поведінку, подібно до різних класів універсальності в теорії критичних явищ [30,31]. Деякі найчастіше спостережувані приклади розподілу ступенів вузлів показані на рис. 9. А саме, розподіл Пуассона:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (2.7)$$

експоненційний розподіл:

$$P(k) \sim e^{-k/\langle k \rangle}, \quad (2.8)$$

степеневий розподіл:

$$P(k) \sim 1/k^\gamma, \quad k \neq 0, \quad \gamma > 0. \quad (2.9)$$

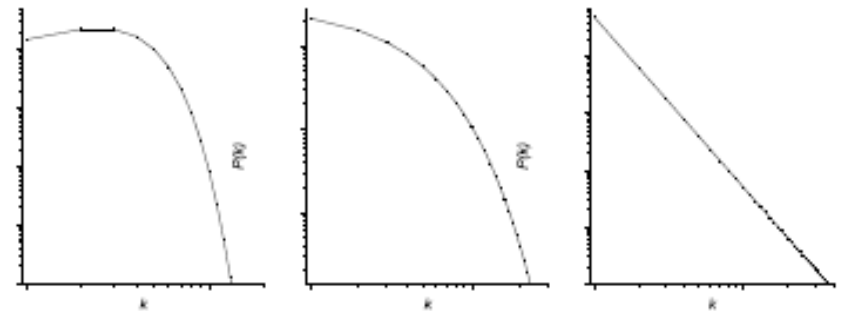


Рис. 9. Розподіл ступенів вузлів  $P(k)$  в log-log масштабі. Зліва направо: розподіл Пуассона  $P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ , експоненційний  $P(k) \sim e^{-k/\langle k \rangle}$  і степеневий  $P(k) \sim 1/k^\gamma$  розподіли.

Попри те, що всі вище наведені функції  $P(k)$  спадають при великих  $k$ , принципова відмінність між розподілами (2.7), (2.8), з одного боку, і розподілом (2.9), з іншого, полягає в тому, що і (2.7), і (2.8) характеризуються певним масштабом. Це положення максимуму для розподілу Пуассона чи характерна довжина спадання для експоненційного розподілу. Водночас степеневий розподіл (2.9) не характеризується типовим масштабом. Мережі зі степеневим розподілом ступенів вузлів (2.9) називають *безмасштабними* (scale-free) мережами. Як буде показано в наступних розділах, саме безмасштабні розподіли часто спостерігаються в складних реально існуючих мережах.

# Математичні моделі мереж – три головних види графів

Складні мережі. Головач Ю., Олемской О., К. фон Фербер та інші // Журнал технічних досліджень т.10, № 4 (2006). 247-289

$$P(k) \sim 1/k^\gamma, \quad k \neq 0, \quad \gamma > 0. \quad (2.9)$$

Інша принципова відмінність цих розподілів полягає в тому, що всі моменти  $P(k)$  існують для розподілів (2.7), (2.8), але не для безмасштабного розподілу (2.9). Справді, для  $P(k)$  (2.9) моменти

$$M_n = \sum_{k=0}^{\infty} k^n P(k) \quad \text{з } n \geq \gamma - 1 \quad (2.10)$$

розбіжні. При степеневому розподілі можливе існування вузлів з дуже високим ступенем (габів, hubs), які практично відсутні в мережах із пуассоновим чи експоненціальним розподілами (2.7), (2.8). Саме наявність габів пояснює поведінку моментів (2.10) і спричиняє багато інших специфічних властивостей безмасштабних мереж.

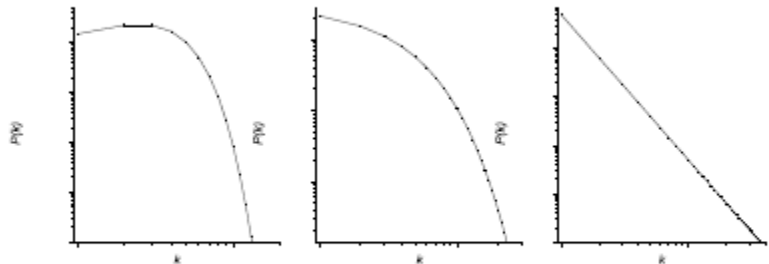


Рис. 9. Розподіл ступенів вузлів  $P(k)$  в log-log масштабі. Зліва направо: розподіл Пуассона  $P(k) = e^{-k} \frac{k^k}{k!}$ , експоненціальний  $P(k) \sim e^{-k/(k)}$  і степеневий  $P(k) \sim 1/k^\gamma$  розподіли.

# Математичні моделі мереж – три головних види графів

*Складні мережі. Головач Ю., Олемской О., К. фон Фербер та інші // Журнал технічних досліджень т.10, № 4 (2006). 247-289*

**А. Класичний випадковий граф Ердоша-Рені  
(мережі реального світу, як правило, не описуються ним)**

# Математичні моделі мереж – три головних види графів

*Складні мережі. Головач Ю., Олемской О., К. фон Фербер та інші // Журнал технічних досліджень т.10, № 4 (2006). 247-289*

**В. Модель Ваттса-Строгатса  
(мережа тісного світу)**

# Математичні моделі мереж – три головних види графів

*Складні мережі. Головач Ю., Олемской О., К. фон Фербер та інші // Журнал технічних досліджень т.10, № 4 (2006). 247-289*

С. Сценарій переважного  
приєднання Барабаші-  
Альберта  
**(безмасштабна зростаюча  
мережа)**

Як СОЦІАЛЬНІ ЗАКЛАДКИ  
(Folksonomy) можуть бути  
змодельовані як **семантика**  
**тегів**, котру можна відкрити ?

# ПОСТАНОВКА ЗАДАЧІ

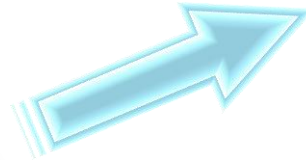
Як СОЦІАЛЬНІ ЗАКЛАДКИ  
(Folksonomy) можуть бути  
змодельовані як **семантика**  
**тегів**, котру можна відкрити ?

Які області знань  
необхідно залучити  
для їх розуміння?

# ПОСТАНОВКА ЗАДАЧІ

Як СОЦІАЛЬНІ ЗАКЛАДКИ  
(Folksonomy) можуть бути  
змодельовані як **семантика**  
**тегів**, котру можна відкрити ?

Які області знань  
необхідно залучити  
для їх розуміння?



Вивчення мережевих  
властивостей структури  
соціальних закладок  
(комплексні безмасштабні  
мережі)



# ПОСТАНОВКА ЗАДАЧІ

Як СОЦІАЛЬНІ ЗАКЛАДКИ  
(Folksonomy) можуть бути  
змодельовані як **семантика**  
**тегів**, котру можна відкрити ?

Які області знань  
необхідно залучити  
для їх розуміння?

Вивчення мережевих  
властивостей структури  
соціальних закладок  
(комплексні безмасштабні  
мережі)

Семантичний  
Web

# ПОСТАНОВКА ЗАДАЧІ

Як СОЦІАЛЬНІ ЗАКЛАДКИ (Folksonomy) можуть бути змодельовані як **семантика тегів**, котру можна відкрити ?

Які області знань необхідно залучити для їх розуміння?

Вивчення мережевих властивостей структури соціальних закладок (комплексні безмасштабні мережі)

Семантичний Web

Подібності між тегами (синоніми, омоніми і т.п.)

# ПОСТАНОВКА ЗАДАЧІ

Як СОЦІАЛЬНІ ЗАКЛАДКИ (Folksonomy) можуть бути змодельовані як **семантика тегів**, котру можна відкрити ?

Які області знань необхідно залучити для їх розуміння?

**ВАШІ ІДЕЇ ?**

Вивчення мережевих властивостей структури соціальних закладок (комплексні безмасштабні мережі)

Семантичний Web

Подібності між тегами (синоніми, омоніми і т.п.)

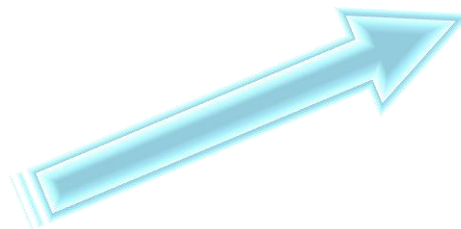
**Соціальні закладки**  
(folksonomy) - що це нам дає ?

**Соціальні закладки**  
(folksonomy) - що це нам дає ?

**ПРАКТИЧНЕ  
ЗАСТОСУВАННЯ**

**Соціальні закладки**  
(folksonomy) - що це нам дає ?

**ПРАКТИЧНЕ  
ЗАСТОСУВАННЯ**



Покращення систем  
рекомендацій  
(**Recommendation Systems**)

**Соціальні закладки**  
(folksonomy) - що це нам дає ?

**ПРАКТИЧНЕ  
ЗАСТОСУВАННЯ**

Покращення систем  
рекомендацій  
(**Recommendation Systems**)

Удосконалення пошуку та  
ранжування  
(**PageRank -> FolkRank**)

**Соціальні закладки**  
(folksonomy) - що це нам дає ?

**ПРАКТИЧНЕ  
ЗАСТОСУВАННЯ**

Покращення систем  
рекомендацій  
(**Recommendation Systems**)

Удосконалення пошуку та  
ранжування  
(**PageRank -> FolkRank**)

Структуризація знань  
семантичними шляхами  
(**DATA Indexing**)



**Соціальні закладки**  
(folksonomy) - що це нам дає ?

**ПРАКТИЧНЕ  
ЗАСТОСУВАННЯ**

Удосконалення  
алгоритмів  
добування даних  
(**DATA Mining**)

Покращення систем  
рекомендацій  
(**Recommendation Systems**)

Удосконалення пошуку та  
ранжування  
(**PageRank -> FolkRank**)

Структуризація знань  
семантичними шляхами  
(**DATA Indexing**)

# Система соціальних букмарків – Del.icio.us

delicious/tags/taltek - Mozilla Firefox

Файл Правка Видгляд Журнал Delicious Закладки Інструменти Довідка

http://delicious.com/tags/taltek

Найбільш відвідвані Початкова сторінка Останні заголовки

Recently Bookmarked Oxford Internet Institut... Thomas Vander Wal: B... Folksonomies - Coope... MIT OpenCourseWare ... Princeton University: ... Free Online Course Ma... Большой город

What's New? Signed in as taltek Inbox Settings Help Sign Out

delicious Home Bookmarks People Tags

Search Delicious Search

Save a new bookmark

taltek's Tags

Bookmarks | Network | Tags | Subscriptions | Inbox

You can also rename tags, delete tags and manage tag bundles.

taltek Type a tag Tags 200

Show all tags Sort: Alphabetically | By size

Access aids AlexeyTitievsky Amazon.com AnD antivirus AR5007EG atheros Automatic.Summarization bibliography bibsonomy Biographic.Information Bluetooth Bookmarks Books BU1-3 C.J.vanRijsbergen C++Builder CAD CADLearning Calvin.Mooers CASE CD-Books ChainProxy Cisco clustering converter Cornell.University cryptography Cutting.Tools CX1 David.Jensen db dblp debian del.icio.us DL Dmitry.Lande Dynamic.Model E-Books e-learning e-prints EndNotes Ergo.Zen.ViewArt ERwin familytree FamilyTreeBuilder flickr flv folksonomies folksonomy for.lifshits free friendfeed full-text Fuzzy.Sets genealogy Gerard.Salton google Google.App.Engine googleapps GoogleCalendar GOST HDD.Player Information.Retrieval Inmarket IP.Tools iPhone iPod IR java jdbc John.Fox Journals journals Laptop Library linguistics Linux logsonomy MakeVideoTutors Making.Names mark.newman MaryKay mediawiki Microlife Microsoft Mitos Mobile MP4Player MS.Office networks newman nlp ODBC Omro OpenSource Paradox PDF photos Protection.Of.Information ProxyCheker python R Ragnarok Referats Ronald.Robert.Yager RX3 sciencedirect.com searchengine semantic Semantic.Nets semantic.web semanticWeb social.bookmarks social.tagging Social.Web socialbookmarking SocialBookmarksManager SocialNetworks Soft software sony statistics Steffen.Staab StuffDipl style.generator Sync tagging Tagging.System TagsEditor TeX thomas.vander.wal tools torrents Toshuba660CDT Transtale UB-201 UDC UML usability Vagif.A.Kasumov vander.wal video Video.Materials VideoTutor WareZ web2.0 wifi wiki wikipedia windows xFolk YouTube Yury.Lifshits Унаучний.портал zotero АРГО аудиокнига БЖД Бібліотеки вадим.зеланд відеолекції Відпочинок вода.питна ву Генеалогія ГОСТ ДискретнийАнализ Дистанційне.Навчання ДСТУ дшлпх Інформаційна.Безпека Книги Конференції.Україна КПК крісло.директорське лекції математика мобільні.телефони Морозильна.Камера навчання нратологія наукаовий.портал нейронні.мережі Нечіткі.Множини Одеса пап Палуша паразити Погода Реферати росія сердечний.моніторинг Соціальна.Мережа статті Струхманчук тонометри трансерфинг.реальности Фото.Альбоми

Зроблено zotero

# Система соціальних букмарків – Del.icio.us

delicious Home Bookmarks People Tags

taltek's Tags  
Bookmarks | Network | Tags | Subscriptions | Inbox

You can also [rename tags](#), [delete tags](#) and [manage tag bundles](#).

taltek Type a tag Tags 200

Show all tags Sort: Alphabetically | By size

Access aids AlexeyTitievsky Amazon.com AnD antivirus AR5007EG atheros Automatic.Summarization  
Biographic.Information Bluetooth Bookmarks Books BU1-3 C.J.vanRijsbergen C++Builder CAD O  
CD-Books ChainProxy Cisco clustering converter Cornell.University cryptography Cutting.Tools C  
**del.icio.us** DL Dmitry.Lande Dynamic.Model E-Books e-learning e-prints EndNotes  
FamilyTreeBuilder flickr flv **folksonomies** folksonomy for.lifshits free friendfeed full-text Fuz  
Google.App.Engine googleapps GoogleCalendar GOST HDD.Player **Information.Retrieval**  
jdbc John.Fox Jomals **journals** Laptop **Library** linguistics Linux logsonomy MakeVideoTutors  
mediawiki Microlife Microsoft Mitos Mobile MP4Player MS.Office networks newman **nlp** ODBC Or  
**Protection.Of.Information** ProxyCheker python R Ragnarok Referats Ronald.Robert.Yager RX3 s  
Semantic.Nets semantic.web semanticWeb **social.bookmarks** social.tagging Social.Web socialbo  
SocialNetworks Soft software sony statistics Steffen.Staab StuffDipl style.generator Sync **tagging** Tag  
thomas.vander.wal tools **torrents** Toshuba660CDT Transtale UB-201 UDC UML usability Vagif.A.Kasumov vander.w  
**VideoTutor** WareZ web2.0 wifi wiki **wikipedia** windows xFolk YouTube Yury.Lifshits Унаучний.портал zotero ЗГО аудиокнига  
БЖД Бібліотеки вадим.зеланд відеолекції Відпочинок вода.питна ву **Генеалогія** ГОСТ ДискретнийАналіз Інформаційне.Навчання  
ДСТУ дшлпх Інформаційна.Безпека Книги Конференції.Україна КПК крісло.директорське лекції математика **мобільні.телефони**  
Морозильна.Камера навчання наратологія **наукаовий.портал** нейронні.мережі Нечіткі.Множини Одеса пап Палуша паразити Погода  
**Реферати** росія сердечний.моніторинг **Соціальна.Мережа** статті Струханчук тонометри трансерфінг.реальности  
Фото.Альбми

Інфохмара (infocloud) тегів для користувача taltek

Зроблено zotero

# Система соціальних букмарків – Del.icio.us

taltek's folksonomies Bookmarks on Delicious - Mozilla Firefox

Файл Правка Видгляд Журнал Delicious Закладки Інструменти Довідка

http://delicious.com/taltek/folksonomies?detail=1

Найбільш відвідвані Початкова сторінка Останні заголовки

delicious Home Bookmarks People Tags

taltek's folksonomies Bookmarks

Bookmarks | Network | Tags | Subscriptions | Inbox

See all folksonomies bookmarks in Popular, Recent, or your Network.

taltek folksonomies Type another tag Bookmarks 23

Sorted by Most Recent

03 SEP 08	Thomas Vander Wal profile	EDIT   DELETE	
	About Gene Smith (Atomiq)	EDIT   DELETE	3
13 AUG 08	Folksonomies - Cooperative Classification and Communication Through Shared Metadata	EDIT   DELETE	2769
23 JUL 08	Folksonomy :: vanderwal.net	EDIT   DELETE	197
21 JUL 08	19. Hypertext 2008: Pittsburgh, PA, USA	EDIT   DELETE	
17 JUL 08	The Community Engine Blog: xFolk: An xhtml microformat for folksonomy	EDIT   DELETE	105
	BibSonomy ::	EDIT   DELETE	1092
	TAGora » Tagging	EDIT   DELETE	
	Stanford InfoLab Publication Server	EDIT   DELETE	46
16 JUL 08	Where the Social Web Meets the Semantic Web	EDIT   DELETE	74

1 2 3 Next >

23 Bookmarks

Save a new bookmark  
Create tag description  
Tag options

Tags

Related Tags 16

- +BibSonomy 1
- +dblp 1
- +folksonomy 1
- +Gene.Smith 1
- +logsonomy 1
- +semanticweb 1
- +social.bookmarks 13
- +Social.Web 1
- +socialbookmarksmanager 1
- +tagging 1
- +Thomas.Vander.Wal 1
- +Vander.Wal 1
- +xFolk 1
- +відеолекції 1
- +наукаовий.портал 1
- +статті 1

Top 10 Tags

All Tags 256

Tag Bundles 2

RSS feed for these Bookmarks © All Rights Reserved

Showing 10 bookmarks per page

Зроблено zotero

# Система соціальних букмарків – Del.icio.us

The screenshot shows a Mozilla Firefox browser window displaying the Del.icio.us profile for 'taltek'. The browser's address bar shows the URL 'http://delicious.com/taltek/folksonomies?detail=1'. The page header includes navigation links like 'Home', 'Bookmarks', 'People', and 'Tags'. The main content area lists bookmarks with their titles, dates, and the number of tags. A blue callout box is overlaid on the page, containing text in Ukrainian. The right sidebar shows a search bar and a list of tags.

**Збережені ресурси сторінок інтернету для користувача taltek для тега "folsonomies" + загальна к-сть посилань на дану сторінку в системі**

Bookmark Title	Date	Number of Tags
Thomas Vander Wal profile	03 SEP 08	3
About Gene Smith (Atomiq)		2769
Folksonomies - Cooperative Classification and Communication Through S	13 AUG 08	197
Folksonomy :: vanderwal.net	23 JUL 08	105
19. Hypertext 2008: Pittsburgh, PA, USA	21 JUL 08	1092
The Community Engine Blog: xFolk: An xhtml microformat for folksonomy e	17 JUL 08	46
BibSonomy ::		74
TAGora » Tagging		
Stanford InfoLab Publication Server		
Where the Social Web Meets the Semantic Web	16 JUL 08	

Navigation: 1 [2] [3] Next > 23 Bookmarks

Footer: RSS feed for these Bookmarks © All Rights Reserved Showing 10 bookmarks per page Зроблено zotero

# Як користувач зберігає цікаві йому ресурси за допомогою del.icio.us

Folksonomy - Wikipedia, the free encyclopedia - Mozilla Firefox

Файл Правка Вигляд Журнал Delicious Закладки Інструменти Довідка

http://en.wikipedia.org/wiki/Folksonomy

Найбільш відвідувані Початкова сторінка Останні заголовки

Recently Bookmarked Курс лекцій з дисцип... Oxford Internet Institut... Thomas Vander Wal: B... Folksonomies - Coope... MIT OpenCourseWare ... Princeton University: ...

taltek's folksonomies Bookmarks o... folksonomies - Google Академія Folksonomy :: vanderwal.net Folksonomy - Entrepreneurship 2.0 Folksonomy - Wikipedia, the free ...

Help us improve Wikipedia by supporting it financially. Log in / create account

article discussion edit this page history

## Folksonomy

From Wikipedia, the free encyclopedia

**Some parts of this article may be misleading.**  
Please help clarify this article. Suggestions may be on the talk page.

*Not to be confused with folk taxonomy.*

**Folksonomy** (also known as **collaborative tagging**, **social classification**, **social indexing**, and **social tagging**) is the practice and method of **collaboratively** creating and managing **tags** to annotate and **categorize content**. In contrast to traditional **subject indexing**, metadata is generated not only by experts but also by creators and consumers of the content. Usually, freely chosen **keywords** are used instead of a **controlled vocabulary**.<sup>[1]</sup> *Folksonomy* is a **portmanteau** of the words *folk* and *taxonomy*, hence a folksonomy is a **user generated** taxonomy.

Folksonomies became popular on the **Web** around 2004 with **social software** applications such as **social bookmarking** or annotating photographs. Websites that support tagging and the principle of folksonomy are referred to in the context of **Web 2.0** because participation is frequently used to visualize the most used tags of a folksonomy. The term folksonomy

Typically, folksonomies are **Internet**-based, although they are also used in other contexts to discover, and navigate over time. A well-developed folksonomy is ideally accessible as an example of websites using folksonomic tagging are **Flickr** and **del.icio.us**, although it

As folksonomies develop in Internet-mediated social environments, users can (general this way, folksonomy users often discover the tag sets of another user who tends to in rewarding gain in the user's capacity to find related content (a practice known as "pivot choice of the search tools that Web sites provide, folksonomies can be seen as a reje

Folksonomy creation and searching tools are not part of the underlying **World Wide Web** level for creating and using tags. These communities are established to enable Web u existing content, such as **Web sites**, books, works in the scientific and scholarly liter

Назад  
Вперед  
Оновити  
Зупинити  
Додати цю сторінку в закладки  
Bookmark This Page In Delicious...  
Зберегти як...  
Відіслати посилання...  
Переглянути зображення тла  
Виділити все  
Перегляд початкового коду сторінки  
Інформація про сторінку  
Властивості

Зроблено

Microsoft Pow... Total Comma... BibSonomy :: ... Folksonomy - ... GateKeeper - c... EN 15:06

# Як користувач зберігає цікаві йому ресурси за допомогою del.icio.us

The screenshot shows a Mozilla Firefox browser window with the address bar displaying `http://en.wikipedia.org/wiki/Folksonomy`. The browser's menu bar includes options like "Файл", "Правка", "Вигляд", "Журнал", "Delicious", "Закладки", "Інструменти", and "Довідка". The browser's toolbar shows navigation buttons and a search bar with the text "folksonomy".

The main content area displays the Wikipedia article for "Folksonomy". The article title is "Folksonomy" and it is identified as "From Wikipedia, the free encyclopedia". A warning box states: "Some parts of this article may be misleading. Please help clarify this article. Suggestions may be on the talk page." The article text includes: "Folksonomy (also known as collaborative tagging tags to annotate and categorize content. In contrast to traditional taxonomies, where tags are freely chosen keywords are used instead of a controlled vocabulary...".

An "Edit Bookmark" dialog box from del.icio.us is overlaid on the article. The dialog box is titled "delicious Edit Bookmark" and shows the following fields: "URL" (http://en.wikipedia.org/wiki/Folksonomy), "TITLE" (Folksonomy - Wikipedia, the free encyclopedia), "KEYWORD" (empty), "NOTES" (empty), and "TAGS" (folksonomies wikipedia web2.0). The dialog box also shows a "Save" button and a "Cancel" button.

The browser's taskbar at the bottom shows several open applications, including "Microsoft Pow...", "Total Comma...", "BibSonomy :: ...", "Folksonomy - ...", and "GateKeeper - C...". The system clock in the bottom right corner shows "15:06".

# Система соціальних букмарків – bibsonomy.org

**BibSonomy :: - Mozilla Firefox**

Файл Правка Видяг Журнал Delicious Закладки Інструменти Довідка

http://www.bibsonomy.org/

Найбільш відвідувані Початкова сторінка Останні заголовки

Recently Bookmarked Oxford Internet Institut... Thomas Vander Wal: B... Folksonomies - Коопе... MIT OpenCourseWare ... Princeton University: ... Free Online Course Ma... Большой город

**BibSonomy ::** search:all <fulltext search here>

A blue social bookmark and publication sharing system.

tags · relations · groups · popular  
myBibSonomy · post bookmark · post publication

logged in as taltek · help · blog · about  
0 picked in basket · edit tags · settings · logout

**bookmarks** RSS XML edit

**i download mp3 here**  
http://alloffmp3.org claims to have a simply enormous catalog - over 200,000 albums online.  
to music Mp3 Download song by cody62086 and 12 other people on Sep 4, 2008, 8:11 PM  
copy

**i download mp3 here**  
http://lavamus.com claims to have a simply enormous catalog - over 200,000 albums online.  
to song music Download Mp3 by cody62086 and 90 other people on Sep 4, 2008, 8:10 PM  
copy

**i download mp3 here**  
http://westsounds.com claims to have a simply enormous catalog - over 200,000 albums online.  
to music song Mp3 Download by cody62086 and 31 other people on Sep 4, 2008, 8:06 PM  
copy

**Rosh Hashanah Gifts**  
to rosh-hashanah-gifts by manishmah on Sep 4, 2008, 8:04 PM  
copy

**publications** RSS BibTeX RDF more edit | pick | unpick

**Haven't we heard this before?**  
Nina / Glick Schiller (2006) Glick Schiller .  
to gg by tss and 1 other person on Sep 4, 2008, 7:41 PM  
pick | copy | URL | BibTeX

**Wer schön sein will, muss leiden : oft ein Nachteil: attraktiv aussehen**  
Rolf Degen Bild der Wissenschaft(6):56-57(2008) – mit Lit.-Hinweisen und Webadresse .  
to Schöner\_Mensch Beurteilung Attraktivität Vorurteil Attraktion by powidl on Sep 4, 2008, 7:25 PM  
pick | copy | BibTeX

**Das schöne Geschlecht – warum die Frau? : Homo sapiens ist die Ausnahme**  
Judith Rauch Bild der Wissenschaft(6):52-55(2008)  
to Tiere Schönheitsideal Geschlechtsunterschied Sexuelle Selektion Mensch  
Evolutionpsychologie by powidl on Sep 4, 2008, 7:10 PM  
pick | copy | URL | BibTeX

**Haven't we heard this before?**  
Nina / Glick Schiller (2006) Glick Schiller .  
to transnationalism by as and 1 other person on Sep 4, 2008, 7:02 PM  
pick | copy | URL | BibTeX

filter:

**- busy tags**  
(alpha | freq) (cloud | list)

2.0 2008 analysis api architecture art assessment babes bbw bdsm belebs berlin big bisexuals bizarre blog blondes boobs book books browser brunettes business butts chrome collaboration community conference css culture data design e-learning eclipse education evaluation food free google graph graphics howto imported industry internet java javascript kommunikation lang:en language learning linux management marketing media model music musik myown network networks news nlp online patterns performance planning problems programming project psychology ranking rdf research science search semantic semanticweb service social software spatial statistics system systems technology test testing theory thesis tools tutorial video visualization web web2.0 wiki wikipedia workshop xml

Зроблено

zotero



# Система соціальних букмарків – bibsonomy.org

BibSonomy :: - Mozilla Firefox

Файл Правка Видгляд Журнал Delicious Закладки Інструменти Довідка

http://www.bibsonomy.org/search/folksonomy

Найбільш відвідувані Початкова сторінка Останні заголовки

Recently Bookmarked folksonomy — CiteSee... SocioPatterns.org Homepage Andreas H... Welcome to ECS EPrin... Functioning Form - Ta... Tag Meaning Disambi... Au Yeung Ching Man, ...

**BibSonomy :: search** ▾ :: folksonomy

A blue social bookmark and publication sharing system.

tags · relations · groups · popular  
myBibSonomy · post bookmark · post publication

logged in as **tatek** · help · blog · about  
0 picked in basket · edit tags · settings · logout

**bookmarks** [RSS](#) [XML](#)

<> edit

**tagCare - take care of your tags**  
Labeling of otherwise related tags  
to folksonomy gardening semantic tools clean tag by hotho on Sep 4, 2008, 12:29 PM  
copy

**Connotea: free online reference management for clinicians and scientists**  
to social reference tags research science folksonomy web bookmarks by kormoran and 48 other people on Aug 11, 2008, 9:42 PM  
copy

**All Tags at Jobazaar**  
to JOBS\_ZZZ\_TO\_SORT\_STUDY+\_WORK by tudp on Aug 10, 2008, 4:05 AM  
copy

**Ontology of Folksonomy**  
to semântica artigos web folksonomia tagging by jacksonmedeiros on Aug 4, 2008, 1:49 PM  
copy

**Analyzing Social Bookmarking Systems: A del.icio.us Cookbook**  
to spam delicious analysis folksonomy by nwdai on Jul 24, 2008, 6:06 PM  
copy

**Tumblr**  
Unlike blogs, tumblelogs aren't designed like a newspaper column. They're the

**publications** [RSS](#) [BibTeX](#) [RDF](#) [more](#)

<> edit | pick | unpick

**Organizing Publications and Bookmarks in BibSonomy**  
Robert Jäschke and Miranda Grahl and Andreas Hotho and Beate Krause and Christoph Schmitz and Gerd Stumme *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007, Banff, Canada, (2007)*  
to bookmarking social bibsonomy by mschuber and 13 other people on Sep 8, 2008, 3:06 PM  
pick | copy | URL | BibTeX

**Logsonomy — A Search Engine Folksonomy**  
Robert Jäschke and Beate Krause and Andreas Hotho and Gerd Stumme *Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM 2008), AAAI Press, (2008)*  
to folksonomy search tripartite tags by mschuber and 5 other people on Sep 8, 2008, 1:49 PM  
pick | copy | URL | BibTeX

**Purpose Tagging - Capturing User Intent to Assist Goal-Oriented Social Search**  
M. Strohmaier *Workshop on Search in Social Media SSM'08, in conjunction with CIKM'08, Napa Valley, USA, (2008)*  
to own web-science networks goals folksonomy commonsense by mstrohm on Sep 7, 2008, 10:19 PM  
pick | copy | URL | BibTeX

**A Folksonomy-Based Model of Web Services for Discovery and Automatic Composition.**  
Eric Bouillet and Mark Febowitz and Hanhua Feng and Zhen Liu and Anand Ranganathan and Anton Riabov *IEEE SCC (1), page389-396. IEEE Computer Society, (2008)*

Зроблено

# Система соціальних букмарків – bibsonomy.org

BibSonomy: bibtex: Organizing Publications and Bookmarks in BibSonomy - Mozilla Firefox

Файл Півка Видгляд Журнал Delicious Закладки Інструменти Довідка

http://www.bibsonomy.org/bibtex/2b3a5e9851647ca0a7dfb62f041872504/mSCHUBER

Найбільш відвідувані Початкова сторінка Останні заголовки

Recently Bookmarked folksonomy — CiteSee... SocioPatterns.org Homepage Andreas H... Welcome to ECS EPrin... Functioning Form - Ta... Tag Meaning Disambi... Au Yeung Ching Man, ...

**BibSonomy :: bibtex** :: Organizing Publications and Bo

A blue social bookmark and publication sharing system.

tags · relations · groups · popular  
myBibSonomy · post bookmark · post publication

logged in as taltek · help · blog · about  
0 picked in basket · edit tags · settings · logout

mschuber's BibTeX entry:

### Organizing Publications and Bookmarks in BibSonomy

Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007, 2007.

Authors: Robert Jäschke and Miranda Grahl and Andreas Hotho and Beate Krause and Christoph Schmitz and Gerd Stumme

Editors: Harith Alani and Natasha Noy and Gerd Stumme and Peter Mika and York Sure and Denny Vrandečić

URL: [http://www2007.org/workshops/paper\\_25.pdf](http://www2007.org/workshops/paper_25.pdf)

Tags: bibsonomy bookmarking social

Abstract: BibSonomy is a web-based social resource sharing system which allows users to organise and share bookmarks and publications in a collaborative manner. Apart from standard folksonomy features such as an intuitive user interface, navigation along all dimensions, or browser integration via RSS feeds, BibSonomy provides tag hierarchies, group management and privacy features, and numerous import and export functions.

pick | copy | URL | BibTeX

```
@inproceedings{paper:jäschke:2007,
  title = {Organizing Publications and Bookmarks in BibSonomy},
  address = {Banff, Canada},
  author = {Robert Jäschke and Miranda Grahl and Andreas Hotho and Beate Krause and Christoph Schmitz and Gerd Stumme},
  booktitle = {Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007},
  editor = {Harith Alani and Natasha Noy and Gerd Stumme and Peter Mika and York Sure and Denny Vrandečić},
  url = {http://www2007.org/workshops/paper_25.pdf},
  year = {2007},
  abstract = {BibSonomy is a web-based social resource sharing system which allows users to organise and share bookmarks and publications in a collaborative manner. Apart from standard folksonomy features such as an intuitive user interface, navigation along all dimensions, or browser integration via RSS feeds, BibSonomy provides tag hierarchies, group management and privacy features, and numerous import and export functions.},
  keywords = {bibsonomy bookmarking social}
}
```

filter:

**- relations**

**- tags**

(alpha | freq) (cloud | list)

agents algorithm analysis application author-topic basic  
bibsonomy blog book bookmarking bug bug-status capture  
centralities challenge classification clustering coefficient  
commonsense concept conceptual conways-law correlation ovs  
dataset defects definitions dependency design development  
directed documents eclipse engineering evaluation  
exploration fix folksonomy forecasting formal framework function  
galois goal goal-extraction goal-oriented goals graph graphs  
hierarchy howto i-star implementation intentional intent  
invention IR jung kernel KM knowledge lab-meeting lattice  
lifetime link-analysis machine-learning management mining  
model msr n-bartite network networks NFR NLP open-source  
organisation paper path patterns personal-development  
planning plot porter predicting programming query ranking RE  
reading-group reasoning recommender refactor relations  
requirement research-group review RG-done search  
semisupervised short similarity small smart social softgoals  
software spam speech spiral statistics stemming story strips  
structure system tag tags taxonomy text to-read tools  
traceability tripartite tutorial types ucinet UML user visualization  
waterfall web web-science web2.0 webservice weighting world  
writing

BibSonomy is offered by the Knowledge and Data Engineering Group of the University of Konstanz, Germany. Contact: [webmaster@bibsonomy.org](mailto:webmaster@bibsonomy.org)

Зроблено

zotero

# Соціальні букмарки: термін **Folksonomy**

Вперше цей термін було запропоновано Томасом Вандер Велом (Thomas Vander Wal) 24 липня 2004 р. у закритому списку розсилки сервера IA Institute. (Під час обговорення такого явища як запуск в 2003 році сайту **Del.icio.us** та згодом **Flickr.com**). Він також відомий як ініціатор терміна ["infocloud"](#) – інфохмара (інформаційна хмара).



<http://www.vanderwal.net/folksonomy.html>

## Creation of Folksonomy Term

On July 23, 2004 in the [IA Institute](#) (then called the Asylomar Institute for Information Architecture (AIFIA)) closed list serve [Gene Smith](#) asked, "Some of you might have noticed services like [Furl](#), [Flickr](#) and [Del.icio.us](#) using user-defined labels or tags to organize and share information.... Is there a name for this kind of informal social classification?". After a few other people answered some other related questions Eric Scheid of [Ironclad Information Architecture](#) responded with "folk classification".

On **July 24, 2004** I responded just after that with, "So the user-created bottom-up categorical structure development with an emergent thesaurus would become a Folksonomy?".



# Соціальні букмарки: термін **Folksonomy**

З серпня 2004 року Джин Сміт (Gene Smith) опублікував цей термін в своєму блозі “[Folksonomy: Social Classification](#)”, таким чином цей термін вийшов за межі закритого списку розсилки сервера AI Institute і став загально відомим.

On August 3, 2004 Gene Smith posted in his blog [Folksonomy: Social Classification](#). This blog post received a lot of traffic and opened up the term folksonomy for others outside the closed IA listserv.

[http://atomiq.org/archives/2004/08/folksonomy\\_social\\_classification.html](http://atomiq.org/archives/2004/08/folksonomy_social_classification.html)

Last week I asked the [AlfIA](#) members' list what they thought about the social classification happening at [Furl](#), [Flickr](#) and [Del.icio.us](#). In each of these systems people classify their pictures/bookmarks/web pages with tags (e.g. [wedding](#)), and then the most popular tags float to the top (e.g. [Flickr's tags](#) or [Del.icio.us](#) on the right). [Thomas Vander Wal](#), in his reply, coined a great name for these informal social categories: **a folksonomy**. I think folksonomies can work well for certain kinds of information because they offer a small reward for using one of the popular categories (such as your photo appearing on a popular page). People who enjoy the social aspects of the system will gravitate to popular categories while still having the freedom to keep their own lists of tags.

...

Still, the idea of socially constructed classification schemes (with no input from an information architect) is interesting. Maybe one of these services will manage to build a social thesaurus.

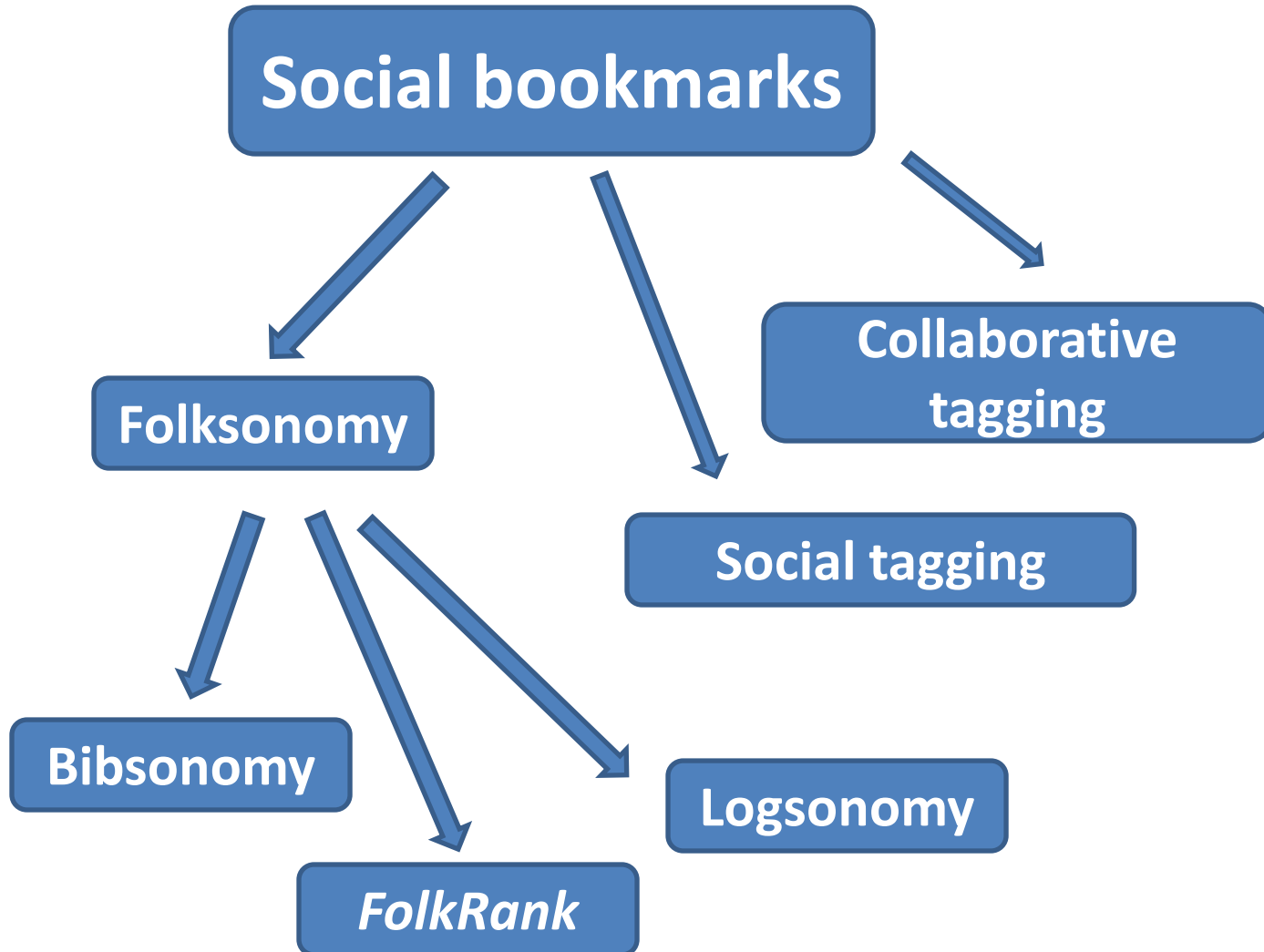
*Posted by Gene Smith on Aug 3, 2004*

# Визначення терміну **Folksonomy**

**Folksonomy** – це спосіб (явище), за допомогою якого люди застосовують до різного роду інформаційних об'єктів (веб-сторінок, фотографій, відео, підкастів і т.п.) свої власні **TAGs** (мітки, характерні слова, ярлики), якими вони оперують, виходячи із власного словникового запасу, маючи на меті легко віднайти необхідну інформацію по “прикріпленому(их)” ярлику(ах).

*“...Folksonomy, a term coined by Thomas Vender Wal in 2005, is a mechanism to describe web resources using people’s own vocabulary.  
Or as defined in Wikipedia (from <http://en.wikipedia.org/wiki/Folksonomy>, accessed 1st April 2007) folksonomy is “... an Internet-based information retrieval methodology consisting of collaboratively generated, open-ended labels that categorize content such as Web pages, online photographs, and Web links.”*

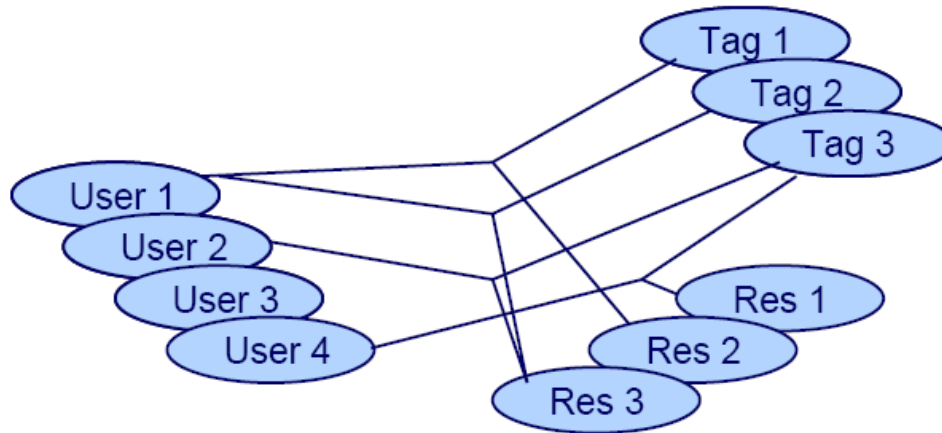
# Інші терміни, які виникли на основі соціальних закладок (Social Bookmarks)



# Інші терміни, які виникли на основі соціальних закладок (Social Bookmarks)

- **BibSonomy** – система, яка дозволяє одночасне використання закладок та елементів BibTeX (бібліографія засобами TeX).  
(Див. статті: 1. “BibSonomy: A Social Bookmark and Publication Sharing System” Andreas Hotho, Robert Jäschke, Christoph Schmitz, Gerd Stumme  
[Knowledge & Data Engineering Group, Department of Mathematics and Computer Science, University of Kassel, Wilhelmshöher Allee 73, D–34121 Kassel, Germany <http://www.kde.cs.uni-kassel.de>, 2 Research Center L3S, Expo Plaza 1, D–30539 Hannover, Germany <http://www.l3s.de>])
- **LogSonomy**. Результатом на запит інформаційно-пошукової системи (ІПС) є список ссилки, що відповідає умовам пошуку. Користувач “клікає” на конкретні лінки (URL) з цього списку, очікуючи, що запропонована інформація буде релевантною (доцільною). “Клікнуті” дані (clickdate) можуть бути представлені як “народна класифікація” (Folksonomy) в цих списках запитів і описана вибраними URL. Результуюча мережева структура, яку автори назвали **LogSonomy**, дуже схожа до Folksonomy.  
*“... Today’s search engines represent the gateway to retrieve information from the World Wide Web. Short queries typically consisting of two to three words describe a user’s information need. In response to the displayed results of the search engine, users click on the links of the result page as they expect the answer to be of relevance. This clickdata can be represented as a folksonomy in which queries are descriptions of clicked URLs. The resulting network structure, which we will term logsonomy is very similar to the one of folksonomies.” [2]*  
(Див. статті:  
1. “Logsonomy—A Search Engine Folksonomy” Robert Jäschke \* ‡, Beate Krause \*, Andreas Hotho \* ‡, Gerd Stumme \* ‡ - 2008  
2. “Logsonomy — Social Information Retrieval With Logdata” Beate Krause \*, Robert Jäschke \* ‡, Andreas Hotho \* ‡, Gerd Stumme \* ‡ – 2008  
[\* Knowledge & Data Engineering Group, University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany, ‡ Research Center L3S, Appelstr. 9a, 30167 Hannover, Germany])

# Математичні основи терміна Folksonomy (1.1)



Трьохчастинний гіперграф (тернерне відношення):

$$F = U \cup T \cup R, \text{ де}$$

- U** – множина користувачів, який створюють запис про даний ресурс;
- T** – множина ключових слів (тегів), яким охарактеризовано ресурси;
- R** – множина ресурсів в Інтернеті (URL)



# Формальная модель Folksonomy

## BibSonomy: A Social Bookmark and Publication Sharing System

Andreas Hotho,<sup>1</sup> Robert Jäschke,<sup>1,2</sup> Christoph Schmitz,<sup>1</sup> Gerd Stumme<sup>1,2</sup>

### A Formal Model for Folksonomies

<sup>1</sup> Knowledge & Data Engineering Group, Department of Mathematics and Computer Science, University of Kassel, Wilhelmshöher Allee 73, D-34121 Kassel, Germany  
<http://www.kde.cs.uni-kassel.de>

<sup>2</sup> Research Center L3S, Expo Plaza 1, D-30539 Hannover, Germany  
<http://www.l3s.de>

A folksonomy describes the users, resources, and tags, and the user-based assignment of tags to resources. We present here a formal definition of folksonomies, which is also underlying our BibSonomy system.

**Definition 1.** A folksonomy is a tuple  $\mathbb{F} := (U, T, R, Y, \prec)$  where

- $U$ ,  $T$ , and  $R$  are finite sets, whose elements are called users, tags and resources, resp.,
- $Y$  is a ternary relation between them, i. e.,  $Y \subseteq U \times T \times R$ , whose elements are called tag assignments (tas for short), and
- $\prec$  is a user-specific subtag/supertag-relation, i. e.,  $\prec \subseteq U \times T \times T$ , called is-a relation.

**Definition 2.** The personomy  $\mathbb{P}_u$  of a given user  $u \in U$  is the restriction of  $\mathbb{F}$  to  $u$ , i. e.,  $\mathbb{P}_u := (T_u, R_u, I_u, \prec_u)$  with  $I_u := \{(t, r) \in T \times R \mid (u, t, r) \in Y\}$ ,  $T_u := \pi_1(I_u)$ ,  $R_u := \pi_2(I_u)$ , and  $\prec_u := \{(t_1, t_2) \in T \times T \mid (u, t_1, t_2) \in \prec\}$ , where  $\pi_i$  denotes the projection on the  $i$ th dimension.

# Формальная модель Folksonomy

Users are typically described by their user ID, and tags may be arbitrary strings. What is considered as a resource depends on the type of system. For instance, in del.icio.us, the resources are URLs, in flickr, the resources are pictures, and in BibSonomy they are either URLs or publication entries. In BibSonomy each resource is represented by a hash value which is described further in Section 4.4.

**Definition 3.** *For convenience we also define the set  $P$  of all posts as*

$$P := \{(u, S, r) \mid u \in U, r \in R, S = \text{tags}(u, r)\}$$

*where, for all  $u \in U$  and  $r \in R$ :  $\text{tags}(u, r) := \{t \in T \mid (u, t, r) \in Y\}$  denotes all tags the user  $u$  assigned to the resource  $r$ .*

If we disregard the is-a relation, we can simply note a folksonomy as a quadruple  $\mathbb{F} := (U, T, R, Y)$ . This structure is known in Formal Concept Analysis [27, 6] as a *triadic context* [13, 26]. An equivalent view on this structure is that of a tripartite (undirected) hypergraph  $G = (V, E)$ , where  $V = U \dot{\cup} T \dot{\cup} R$  is the set of nodes, and  $E = \{\{u, t, r\} \mid (u, t, r) \in Y\}$  is the set of hyperedges.

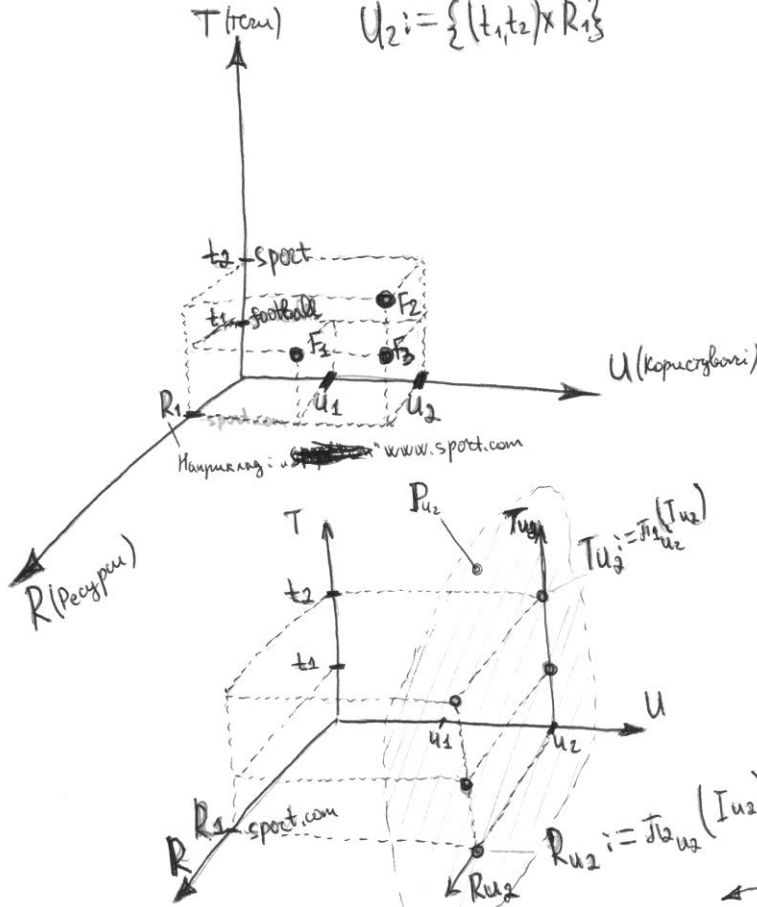
In a typical folksonomy system every tag assignment is connected with several other properties like date, group or resource type. For sake of simplicity we disregard these properties for the rest of the paper, unless stated otherwise.

# Формальна модель Folksonomy

Bibsonomy: A Social Bookmark and Publication Sharing System (Andreas Hotho, Robert Jäschke, Christoph Schmitz, Gerd Stumme)

$$U_1 := \{(t_1) \times R_1\}$$

$$U_2 := \{(t_1, t_2) \times R_1\}$$



Визначення 1:  $F$  - фолксомія

$$F := (U, T, R, Y, \prec)$$

$U, T, R$  - множини Users, Tags, Resources

$Y$  - тернерне відношення:  $Y \subseteq U \times T \times R$

елементи якого назвемо tag assignments (або простоти)

$\prec$  - запит користувачами tag/над тегами - зв'язки

$\prec \subseteq U \times T \times T$  - назвемо їх зв'язки

Визначення 2:  $P_u$  - персоналія (персоналізація) користувача  $u \in U$   
 є відношення для  $F$  по користувачу (або в площині користувача)

$$P_u := (T_u, R_u, I_u, \prec_u), \text{ де } I_u := \{(t, z) \in T \times R \mid (u, t, z) \in Y\}$$

$$T_u := \pi_1(I_u), R_u := \pi_2(I_u), \prec_u := \{(t_1, t_2) \in T \times T \mid (u, t_1, t_2) \in \prec\}$$

де  $\pi_i$  - означає проєкцію в  $i$ -му виміру (вісь  $T_u$ , вісь  $R_u$  проєкції)

Користувачі  $U$  описані їхніми ID, Тегів Tags - збіркою стрічок  
 Для прикладу:  
 delicious - URIs  
 flickr - Pictures  
 Bibsonomy - адреса на публікацію/передавання де знається теги.

FCA [27, 6]  
 Triadic context [13, 26]

Визначення 3: Два зручності ми також задаємо множини  $P$  всіх можлив:  
 $P := \{(u, S, z) \mid u \in U, z \in R, S = \text{tags}(u, z), \text{де } \text{tags}(u, z) \text{ є набір всіх } t \in U$   
 $i \in R: \text{tags}(u, z) := \{t \in T \mid (u, t, z) \in Y\}$  вказує на всі теги  
 користувача  $u$ , що описав ресурс  $z$ .

Якщо ми знехтуємо що є зв'язки, то ми можемо просто описати фолксомію 4-арне (quadruple)  $F' := (U, T, R, Y)$

Ця структура є відомою з Formal Concept Analysis [27, 6]  
 як "Triadic context" [13, 26]. Еквівалентне зображення цієї структури  
 є тернерний (ненаправлений) гіперграф  $G = (V, E)$ , де  
 $V = U \cup T \cup R$  є множиною вершин;  $E = \{(u, t, z) \mid (u, t, z) \in Y\}$  є гіперрефер.

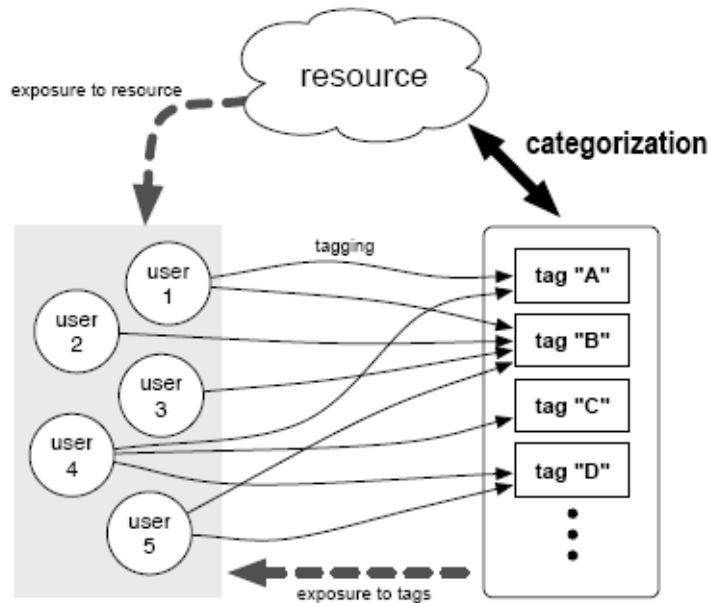
# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by **Ciro Cattuto**, 2006 (2008)

### Collaborative Tagging and Semiotic Dynamics

Ciro Cattuto\*

Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
Compendio Viminale, 00184 Roma, Italy and  
Dipartimento di Fisica, Università di Roma "La Sapienza"  
P.le A. Moro, 2, 00185 Roma, Italy



Collaborative tagging has been quickly gaining ground because of its ability to recruit the activity of web users into effectively organizing and sharing vast amounts of information. Here we collect data from a popular system and investigate the statistical properties of tag co-occurrence. We introduce a stochastic model of user behavior embodying two main aspects of collaborative tagging: (i) a frequency-bias mechanism related to the idea that users are exposed to each other's tagging activity; (ii) a notion of memory – or aging of resources – in the form of a heavy-tailed access to the past state of the system. Remarkably, our simple modeling is able to account quantitatively for the observed experimental features, with a surprisingly high accuracy. This points in the direction of a universal behavior of users, who – despite the complexity of their own cognitive processes and the uncoordinated and selfish nature of their tagging activity – appear to follow simple activity patterns.

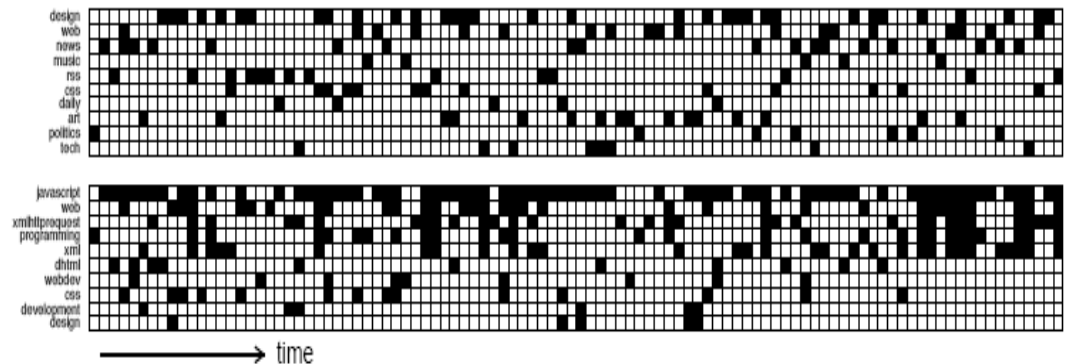


FIG. 2: Tagging activity: a time-ordered sequence of tagging events is graphically rendered by marking the tags co-occurring with *blog* (top panel) or *ajax* (bottom panel) in an experimental sequence of posts on *del.icio.us*. In each panel, columns represent single tagging events (posts) and rows correspond to the 10 most frequent tags co-occurring with either *blog* (top panel) or *ajax* (bottom panel). 100 tagging events are shown in each panel, temporally ordered from left to right. Only posts involving at least one of the 10 top-ranked tags are shown. For each tagging event (column), a filled cell marks the presence of the tag in the corresponding row, while an empty cell indicates its absence. A qualitative difference between *blog* (top panel) and *ajax* (bottom panel) is clearly visible, where a higher density at low-rank tags characterizes the semantically narrower *ajax* term. This corresponds to the steeper low-rank behavior observed in the frequency-rank plot for *ajax* (Fig. 3).

FIG. 1: Schematic depiction of the collaborative tagging process: web users are exposed to a resource and freely associate tags with it. Their interaction with the system also exposes them to tags previously entered by themselves and by other users. The aggregated activity of users leads to an emergent categorization of resources in terms of tags shared by a community.

# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by **Ciro Cattuto**, 2006 (2008)

Collaborative Tagging and Semiotic Dynamics

Ciro Cattuto

Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
Compendio Viminale, 00184 Roma, Italy and  
Dipartimento di Fisica, Università di Roma "La Sapienza"  
P.le A. Moro, 2, 00185 Roma, Italy

## A YULE-SIMON'S MODEL WITH LONG-TERM MEMORY

We now aim at gaining a deeper insight into the phenomenology reported above. In order to model the observed frequency-rank behavior for the full range of ranking values, we introduce a new version of the "rich-get-richer" Yule-Simon's stochastic model [13, 14] by enhancing it with a fat-tailed memory kernel.

We tested this hypothesis with real data extracted from *del.icio.us*: Fig. 4 shows the temporal auto-correlation function for the sequence of tags co-occurring with *blog*. Such a sequence is constructed by consecutively appending the tags associated with each post, respecting the temporal order of posts. Correlations are computed inside three consecutive windows of length  $T$ , starting at different times  $t_w$ ,

$$C(\Delta t, t_w) = \frac{1}{T - \Delta t} \sum_{t=t_w+1}^{t=t_w+T-\Delta t} \delta(\text{tag}(t + \Delta t), \text{tag}(t)),$$

where  $\delta(\text{tag}(t + \Delta t), \text{tag}(t))$  is the usual Kronecker delta function, taking the value 1 when the same tag occurs at times  $t$  and  $t + \Delta t$ . From Fig. 4 it is apparent that the cor-

relation function is non-stationary over time. Moreover, for each value of the initial time  $t_w$  a power-law behavior is observed:  $C(\Delta t, t_w) = a(t_w)/(\Delta t + \delta(t_w)) + c(t_w)$ , where  $a(t_w)$  is a time-dependent normalization factor and  $\delta(t_w)$  is a phenomenological time scale, slowly increasing with the "age"  $t_w$  of the system.  $c(t_w)$  is the correlation that one would expect in a random sequence of tags distributed according to the frequency-rank distribution  $P_{T, t_w}(R)$  pertaining to the relevant data window. Denoting by  $R = R_{\max}(T, t_w)$  the number of distinct tags occurring in the window  $[t_w, t_w + T]$ , we have  $c(t_w) = \sum_{R=1}^{R=R_{\max}(T, t_w)} P_{T, t_w}^2(R)$ .

Our modification of the Yule-Simon's model thus consists in weighting the probability of choosing an existing word (tag) according to a power-law kernel. This hypothesis about the functional form of the memory kernel is also supported by findings in Cognitive Psychology [17], where power laws of latency and frequency have been shown to model human memory.

# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by **Ciro Cattuto**, 2006 (2008)

Collaborative Tagging and Semiotic Dynamics

Ciro Cattuto

Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
Compendio Viminale, 00184 Roma, Italy and  
Dipartimento di Fisica, Università di Roma "La Sapienza"  
P.le A. Moro, 2, 00185 Roma, Italy

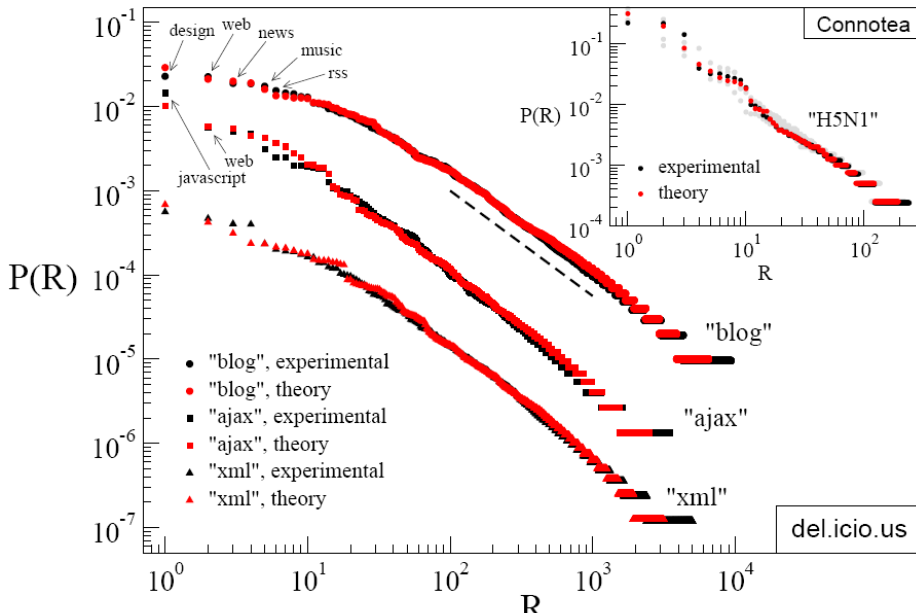


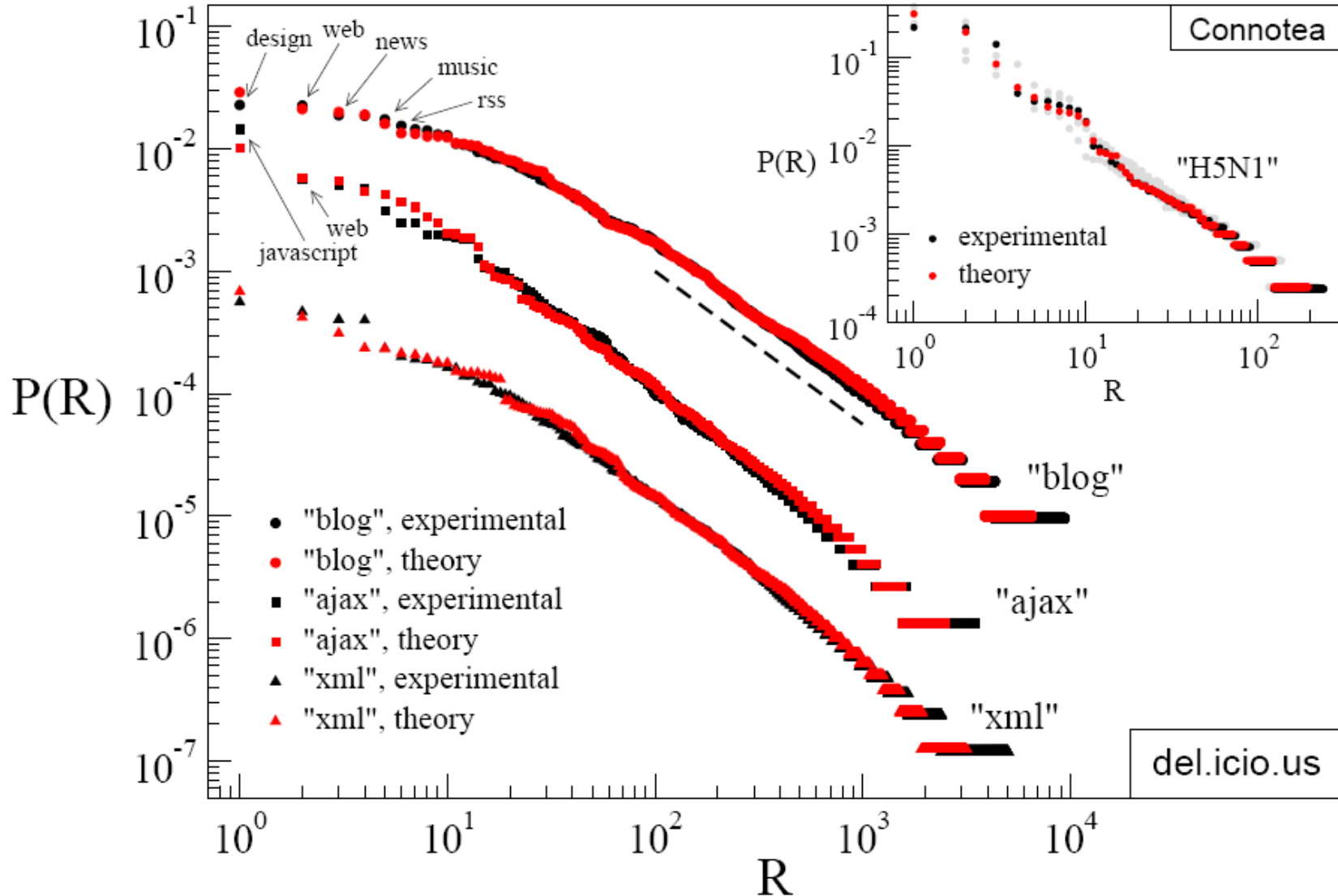
FIG. 3: Frequency-rank plots for tags co-occurring with a selected tag: experimental data (black symbols) are shown for *del.icio.us* (circles for tags co-occurring with the popular tag *blog*, squares for *ajax* and triangles for *xml*) and *Connotea* (inset, black circles for the *H5N1* tag). For the sake of clarity, the curves for *ajax* and *xml* are shifted down by one and two decades, respectively. Details about the experimental datasets are reported in Table I. All curves exhibit a power-law decay for high ranks (a dashed line corresponding to the power law  $R^{-5/4}$  is provided as an aid for eye) and a shallower behavior for low ranks. To make contact with Fig. 2, some of the highest-frequency tags co-occurring with *blog* and *ajax* are explicitly indicated with arrows. Red symbols are theoretical data obtained by computer simulation of the stochastic process described in the text (Fig. 5). The parameters of the model, i.e. the probability  $p$ , the memory parameter  $\tau$  and the initial number of words  $n_0$  were adjusted to match the experimental data, giving approximately  $p = 0.06$ ,  $\tau = 100$  and  $n_0 = 100$  for *blog*,  $p = 0.03$ ,  $\tau = 20$  and  $n_0 = 50$  for *ajax*, and  $p = 0.034$ ,  $\tau = 40$  and  $n_0 = 110$  for *xml*. Inset: *Connotea* is a much younger system than *del.icio.us* and the corresponding dataset is smaller and noisier. Nevertheless, a good match with experimental data can be obtained for  $p = 0.05$ ,  $\tau = 120$  and  $n_0 = 7$  (red circles), demonstrating that our model also applies to the early stages of development of a folksonomy. Gray circles correspond to different realizations of the simulated dynamics.

# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by **Ciro Cattuto**, 2006 (2008)

Collaborative Tagging and Semiotic Dynamics


Ciro Cattuto  
Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
Compendio Viminale, 00184 Roma, Italy and  
Dipartimento di Fisica, Università di Roma "La Sapienza"  
P.le A. Moro, 2, 00185 Roma, Italy



# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by **Ciro Cattuto**, 2006 (2008)

Collaborative Tagging and Semiotic Dynamics

Ciro Cattuto   
 Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
 Compendio Viminale, 00184 Roma, Italy and  
 Dipartimento di Fisica, Università di Roma "La Sapienza"  
 P.le A. Moro 2, 00185 Roma, Italy

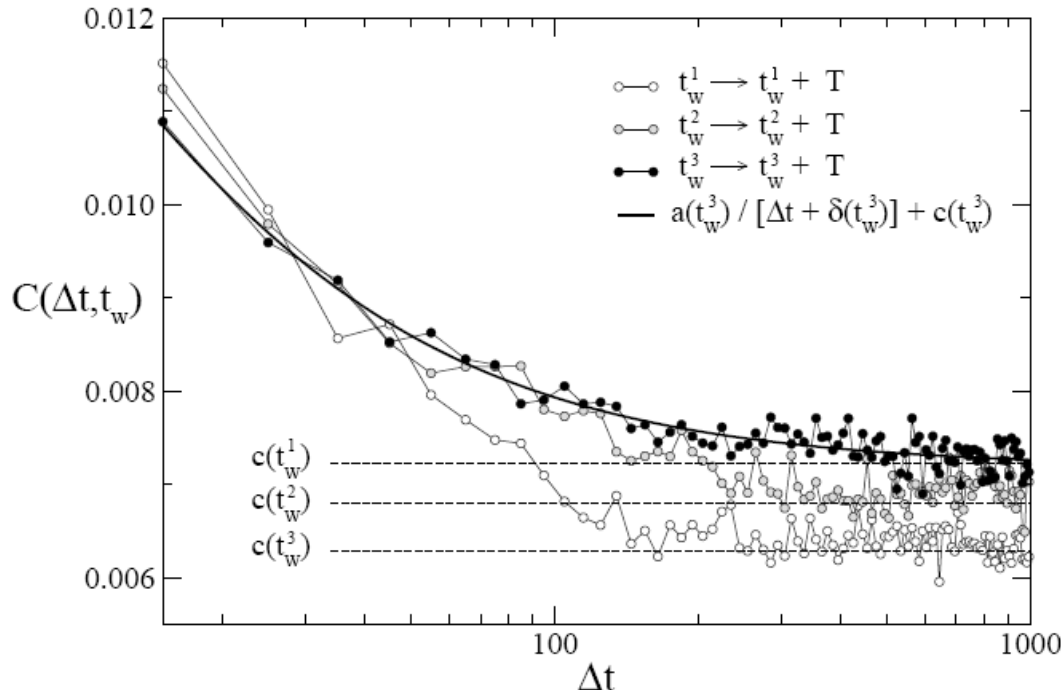


FIG. 4: Tag-tag correlation functions and non-stationarity. The tag-tag correlation function  $C(\Delta t, t_w)$  is computed over three consecutive and equally long ( $T = 30000$  tags each) subsets of the *blog* dataset, starting respectively at positions  $t_w^1 = 10000$ ,  $t_w^2 = 40000$  and  $t_w^3 = 70000$  within the collected sequence. Short-range correlations are clearly visible, slowly decaying towards a long-range plateau value. The non-stationary character of correlations is visible both at short range, where the value of the correlation function decays with  $t_w$ , and at long range, where the asymptotic correlation increases with  $t_w$ . The long-range correlations (dashed lines) can be estimated as the natural correlation present in a random sequence containing a finite number of tags: on using the appropriate ranked distribution of tag frequencies within each window (see text) the values  $c(t_w^1)$ ,  $c(t_w^2)$  and  $c(t_w^3)$  can be computed, matching the measured plateau of the correlation functions. The thick line is a fit to the fat-tailed memory kernel described in the text.

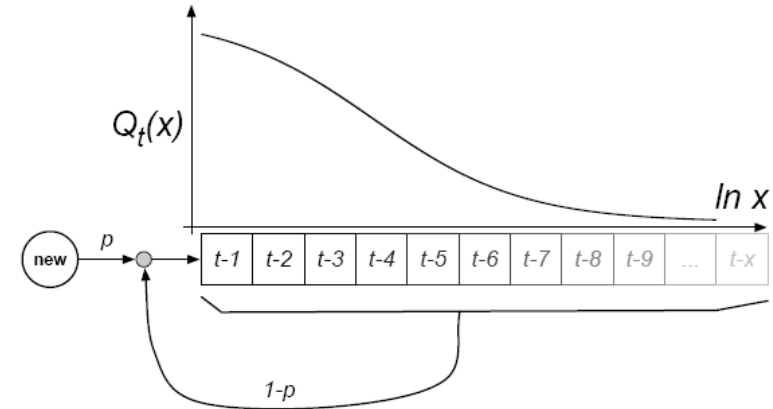


FIG. 5: A Yule-Simon's process with long-term memory. A synthetic stream of tags is generated by iterating the following step: with probability  $p$  a new tag is created and appended to the stream, while with probability  $1 - p$  a tag is copied from the past of the stream and appended to it. The probability of selecting a tag located  $x$  steps into the past is given by the long-range memory kernel  $Q_t(x)$ , which provides a fat-tailed access to the past of the stream.



# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by **Ciro Cattuto**, 2006 (2008)

Collaborative Tagging and Semiotic Dynamics

Ciro Cattuto<sup>1</sup>  
 Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
 Compendio Viminale, 00184 Roma, Italy and

	design	web	news	music	res	css	daily	art	politics	tech	technology	blogs	software	media	programming
inspiration	<b>0.392</b>	0.199	0.028	0.011	<b>0.000</b>	0.208	0.011	0.132	<b>0.000</b>	0.011	0.021	0.032	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
socialsoftware	0.086	0.242	0.022	0.011	0.099	0.044	<b>0.000</b>	0.022	<b>0.000</b>	0.055	0.055	0.121	0.176	0.077	0.011
economics	<b>0.000</b>	0.119	0.090	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.119	<b>0.000</b>	<b>0.552</b>	0.030	0.045	0.030	<b>0.000</b>	<b>0.015</b>	<b>0.000</b>
opensource	0.081	0.144	0.108	0.018	0.063	0.009	0.027	0.009	<b>0.000</b>	0.054	0.045	0.090	0.243	0.036	0.072
computer	0.167	0.080	0.093	0.013	0.027	0.073	0.033	0.020	0.020	0.080	0.133	0.080	0.127	0.007	0.087
python	0.031	0.138	0.077	<b>0.000</b>	0.092	0.046	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.015	0.046	<b>0.000</b>	0.092	<b>0.000</b>	<b>0.462</b>
tagging	0.234	0.213	0.021	<b>0.000</b>	0.085	0.085	0.021	<b>0.000</b>	0.021	0.021	0.085	0.043	0.170	<b>0.000</b>	<b>0.000</b>
comics	0.089	0.089	0.089	0.054	<b>0.000</b>	0.018	0.054	<b>0.482</b>	0.018	<b>0.000</b>	0.036	0.054	<b>0.000</b>	0.018	<b>0.000</b>
research	0.096	0.135	0.096	<b>0.000</b>	0.058	<b>0.000</b>	0.058	0.038	0.038	0.019	0.154	0.077	0.096	0.096	0.038
law	0.012	0.061	0.134	0.037	<b>0.000</b>	<b>0.000</b>	0.110	<b>0.000</b>	<b>0.378</b>	0.024	0.195	0.049	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
hack	0.080	0.080	0.089	0.034	0.057	0.046	0.126	0.023	0.034	0.161	0.103	0.057	0.089	0.023	0.034
xhtml	<b>0.311</b>	<b>0.282</b>	0.016	0.004	0.016	<b>0.336</b>	<b>0.000</b>	0.004	<b>0.000</b>	0.008	0.012	0.004	0.012	0.004	0.008
humour	0.125	0.042	0.125	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.125	0.208	0.063	0.063	0.042	<b>0.000</b>	<b>0.000</b>	0.042	0.125
management	0.147	0.191	<b>0.000</b>	<b>0.000</b>	0.029	0.015	0.118	0.015	<b>0.000</b>	0.015	0.103	0.103	0.162	0.029	0.074
movies	0.080	0.080	0.240	0.120	<b>0.000</b>	<b>0.000</b>	0.100	0.140	0.080	<b>0.000</b>	0.020	0.100	<b>0.000</b>	0.040	<b>0.000</b>
diy	0.151	0.081	0.047	0.023	<b>0.000</b>	<b>0.000</b>	0.081	0.186	0.012	0.198	0.093	0.047	0.058	0.012	0.012
life	0.231	<b>0.000</b>	<b>0.000</b>	0.019	<b>0.000</b>	<b>0.000</b>	0.096	0.250	0.058	0.058	0.250	<b>0.000</b>	0.019	<b>0.000</b>	0.019
tag	0.125	<b>0.354</b>	<b>0.000</b>	<b>0.000</b>	0.250	0.021	<b>0.000</b>	0.021	<b>0.000</b>	0.021	0.062	0.021	0.063	0.021	0.021
maps	0.200	0.150	0.050	<b>0.000</b>	0.050	<b>0.000</b>	<b>0.000</b>	0.050	0.050	0.250	0.050	0.050	0.050	0.050	<b>0.000</b>
ideas	0.167	0.103	0.090	<b>0.000</b>	0.026	0.077	0.077	0.051	0.038	0.064	0.103	0.077	0.051	0.064	0.013
architecture	<b>0.528</b>	0.063	0.074	0.011	<b>0.000</b>	<b>0.000</b>	0.095	0.137	<b>0.000</b>	<b>0.000</b>	0.032	0.011	0.011	<b>0.000</b>	0.042
organization	<b>0.000</b>	0.042	<b>0.000</b>	0.021	<b>0.000</b>	<b>0.000</b>	<b>0.479</b>	0.021	<b>0.000</b>	0.062	0.125	0.021	0.146	0.042	0.042
plugin	0.156	<b>0.281</b>	<b>0.000</b>	0.031	0.031	0.031	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.031	0.031	<b>0.312</b>	0.031	0.062
blogroll	0.054	0.089	0.107	0.071	0.054	<b>0.000</b>	0.054	0.054	0.071	0.089	0.107	0.125	0.054	0.018	0.054
information	0.065	<b>0.280</b>	0.172	<b>0.000</b>	0.097	0.022	0.022	0.011	0.043	0.054	0.075	0.054	0.054	0.032	0.022
articles	0.159	0.136	0.091	0.011	<b>0.000</b>	0.068	0.057	0.011	0.091	0.057	0.080	0.045	0.091	0.045	0.057
resource	<b>0.267</b>	0.202	0.032	0.032	0.032	0.191	0.021	0.032	0.011	0.053	0.011	0.011	0.032	<b>0.000</b>	0.053
illustration	<b>0.339</b>	0.104	0.009	0.028	<b>0.000</b>	0.078	0.017	<b>0.400</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.017	0.009	<b>0.000</b>	<b>0.000</b>
mobile	0.021	0.085	0.108	0.021	0.021	<b>0.000</b>	0.128	0.108	<b>0.000</b>	0.108	<b>0.255</b>	0.043	0.064	0.043	<b>0.000</b>
liberal	<b>0.000</b>	<b>0.000</b>	0.125	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.087	<b>0.000</b>	<b>0.692</b>	<b>0.000</b>	<b>0.000</b>	0.058	<b>0.000</b>	0.058	<b>0.000</b>

FIG. 6: Co-occurrence table: columns correspond to the 15 top-ranked tags co-occurring with *blog*, in descending order of frequency from left to right. Rows correspond to 30 low-frequency tags co-occurring with *blog* (frequencies ranking between 100th and 200th). Each row is a normalized co-occurrence histogram representing a “categorization” of the corresponding tag in terms of the top-ranked tags. Numbers in red (bold face) denote co-occurrence probabilities in excess of 25%. Zeros (no co-occurrence) are marked in blue (bold face).

# Дослідження статистичних властивостей, та стохастична модель поведінки користувача для систем з тегами

## by [Ciro Cattuto](#), 2006 (2008)

Collaborative Tagging and Semiotic Dynamics

[Ciro Cattuto](#)  
Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
Compendio Viminale, 00184 Roma, Italy and  
Dipartimento di Fisica, Università di Roma "La Sapienza"  
P.le A. Moro, 2, 00185 Roma, Italy

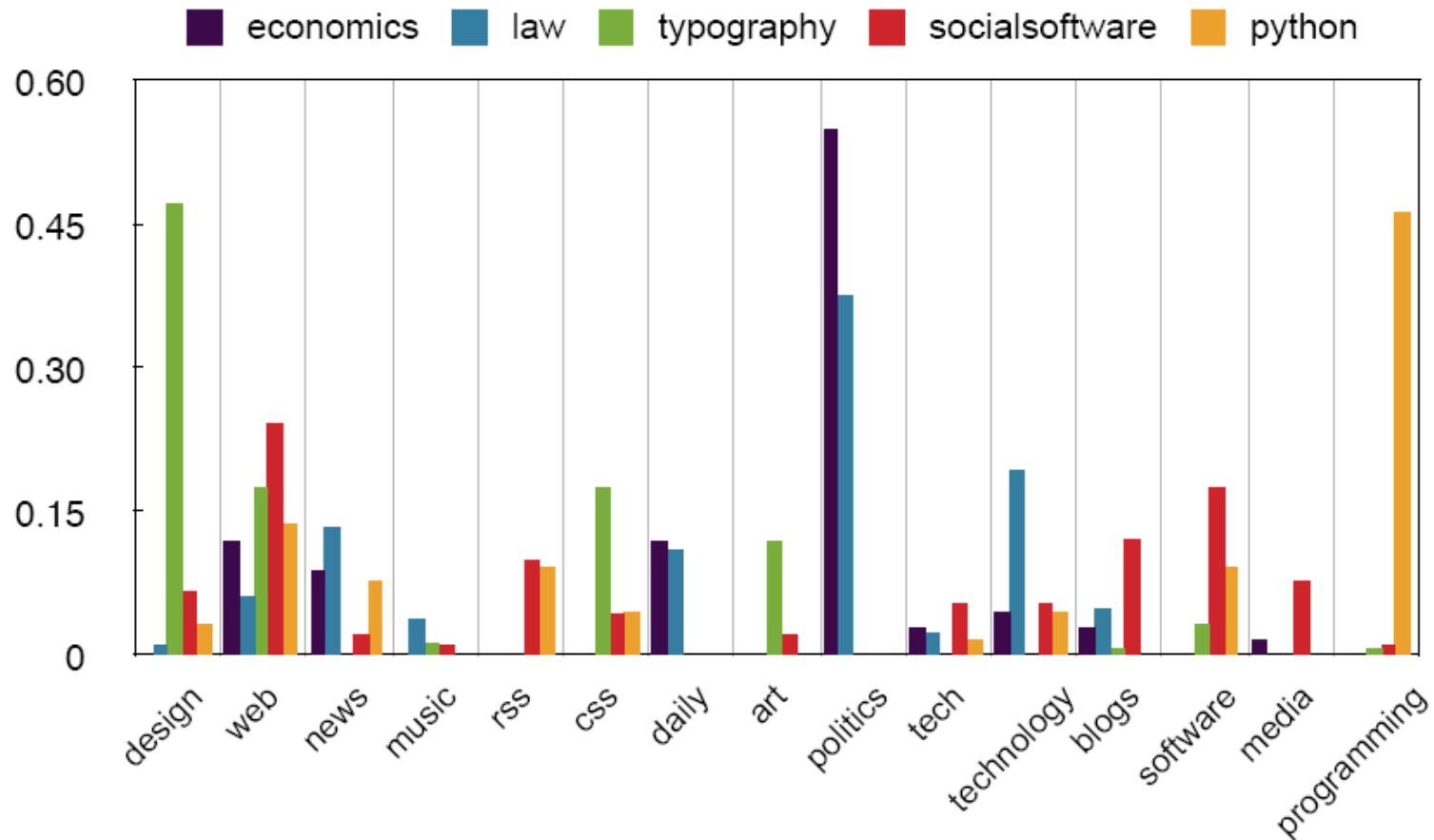


FIG. 7: Co-occurrence patterns for 5 of the low-frequency (high-rank) tags of Fig. 6 (see legend at the top). The colored bars display the “fingerprint” of the selected tags in terms of their co-occurrence with the 15 top-ranked tags (the same ones reported in the top row of Fig. 6).

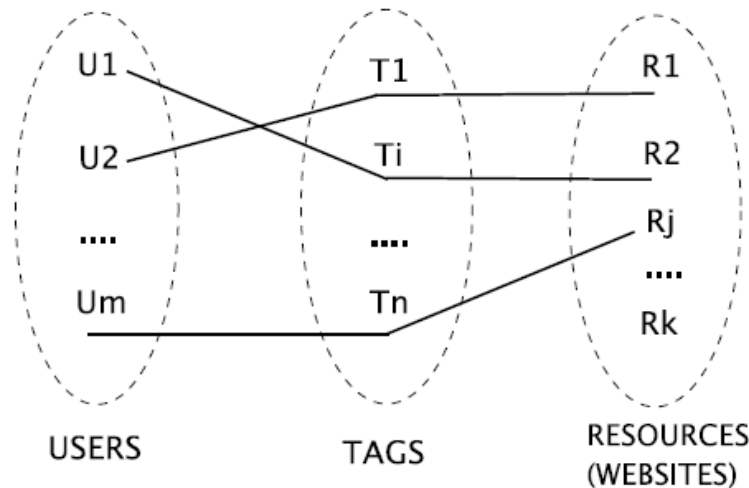
# Аналіз динаміки і семантики tagging system by Harry Halpin, Valentin Robu, Hana Shepherd, (SAAW'06), 2006

## The Dynamics and Semantics of Collaborative Tagging

Harry Halpin  
University of Edinburgh  
2 Buccleuch Place  
Edinburgh, Scotland  
H.Halpin@ed.ac.uk

Valentin Robu  
Dutch Center for Mathematics  
and Computer Science  
Kruislaan 413  
Amsterdam, Netherlands  
robu@cwi.nl

Hana Shepherd  
Princeton University  
Wallace Hall  
Princeton, NJ USA  
hshepher@princeton.edu



There are three main entities that compose any tagging system:

- The users of the system (people who actually do the tagging)
- The tags themselves
- The resources being tagged (in this case, the websites)

Each of these can be seen as forming separate spaces consisting of sets of vertices, which are linked together by edges (see Fig. 1). The first space, the *user space*, consists of the set of all users of the tagging system, where each vertex is a user. The second space is the *tag space*, the set of all tags, where a tag corresponds to a term (“music”) or neologism (“toread”) in natural language. The third space is the *resource space*, the set of all resources, where each resource is normally denoted by a unique URI.<sup>1</sup> A tagging instance can be seen as the two edges that links together a user to a tag and then that tag to a given website or resource. Note that a tagging instance can associate a date with its tuple of a user, a tag(s), and a resource.

Figure 1: Tripartite graph structure of a tagging system. An edge linking a user, a tag and a resource (website) represents one tagging instance

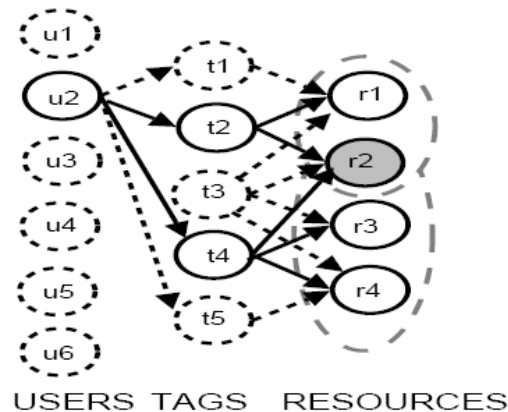


Figure 3: Tripartite tagging system graph used for search. The dotted edges represent options, while the dark edges represent a particular user engaging in a search for the shaded resource

# Трѳохчастинна (tripartite) мѳрежа folksonomy

## by R. Lambiotte, M. Ausloos, 2005

Collaborative tagging as a tripartite network

R. Lambiotte<sup>1</sup> and M. Ausloos<sup>1</sup>

*SUPRATECS, Universitѳ de Liѳge,*

*B5 Sart-Tilman, B-4000 Liѳge, Belgium*

(Dated: 09/07/2005)

Abstract

We describe online collaborative communities by tripartite networks, the nodes being persons, items and tags. We introduce projection methods in order to uncover the structures of the networks, i.e. communities of users, genre families... To do so, we focus on the correlations between the nodes, depending on their profiles, and use percolation techniques that consist in removing less correlated links and observing the shaping of disconnected islands. The structuring of the network is visualised by using a tree representation. The notion of diversity in the system is also discussed.

Tripartite structure

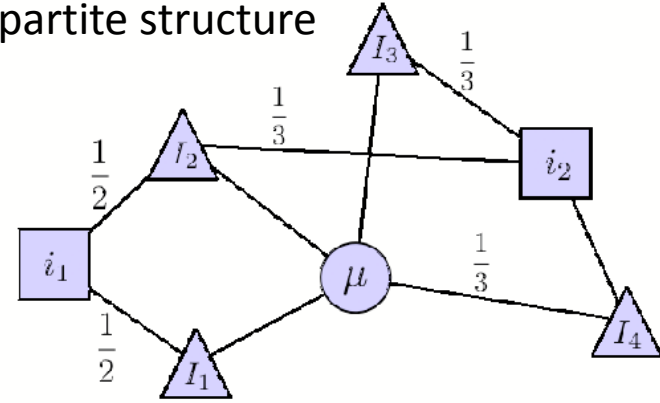


FIG. 1: Tripartite structure of the tagging system. In this example, user  $\mu$  owns two items  $i_1$  and  $i_2$  that are respectively tagged by two keywords,  $(I_1, I_2)$ , and three keywords  $(I_2, I_3, I_4)$ .

matrix  $\overline{\overline{\sigma}}^\mu$ :

$$\overline{\overline{\sigma}}^\mu = \begin{pmatrix} 0 & \dots & 1/2 & \dots & 1/2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & 1/3 & \dots & 1/3 & \dots & \dots & 1/3 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \quad (1)$$

where  $\overline{\overline{\sigma}}_{iI}^\mu$  denotes the weight of tag  $I$  in its description of  $i$ , so that  $\sum_I \overline{\overline{\sigma}}_{iI}^\mu = 1$  if  $\mu$  owns  $i$  and zero otherwise. Each item and each tag is also characterised by similar matrices that we note  $\overline{\overline{\gamma}}^i$  and  $\overline{\overline{\alpha}}^I$  respectively.

## Visualizing tags over time

1) May 2006 in: *Proceedings of the 15th International WWW Conference*, (10pages)

2) August 2007 *ACM Transactions on the Web (TWEB)*, Volume 1 Issue 2 (22pages)

# В статті представлення візуалізація відхилень(тенденцій) (trands) фолксономії

## 4.1 Interestingness

We now define what we mean by “interesting.” First, we require a little notation. Let  $0, \dots, T - 1$  be discrete points in time (also called timestamps) and let  $U = \{u_1, \dots\}$  be the universe of objects — in our case, objects are tags. An object  $u \in U$  has a multiset of timestamps associated with it, indicating its occurrence over time. Because the occurrences are a multiset, an object may occur many times during the same timestep. Let  $\gamma(u, t)$  denote the number of times the object  $u$  occurs at time  $t$ . Let  $\gamma(u) = \sum_{t=0}^{T-1} \gamma(u, t)$  denote the total number of occurrences of object  $u$ .

Our measure of interestingness should have the following properties:

1. An object should be considered more interesting during a particular interval if it occurs more frequently within the interval, and less frequently outside the interval.
2. A highly infrequent object that happens to occur only during a particular time interval should not necessarily be the most interesting object for that time interval.

Let  $I = [a, b]$  be a time interval, where  $0 \leq a < b \leq T$ . We introduce a measure to meet the tradoffs implied by these conditions. As a baseline, we consider the probability that a particular object occurs within  $I$ : this meets both aspects of Property 1 above. We modify this baseline idea to introduce a regularization constant  $C$ , a positive integer, to meet the requirements of Property 2. For any object  $u$  and interval  $I$ , our measure of the interestingness of  $u$  during  $I$  is given by:

$$\text{Int}(u, I) = \sum_{t \in I} \gamma(u, t) / (C + \gamma(u)). \quad (1)$$

This quantity, reminiscent of ‘tf-idf’ in information retrieval (see also [15]), is aimed at capturing the most interesting — not necessarily the most frequent — objects that occur in the interval  $I$ . The parameter  $C$  ensures that objects that occur only in  $I$  but very few number of times do not take advantage of the ‘idf’ term in the definition. Finally, the most interesting objects for  $I$  are those with the highest values of  $\text{Int}(\cdot, I)$ , with the actual value measuring the “intensity” of the interestingness.

Our definition of interestingness is intentionally simple; one could imagine more sophisticated definitions. Note that our definition of interestingness is linear: if  $I_1$  and  $I_2$  are disjoint intervals, then  $\text{Int}(u, I_1 \cup I_2) = \text{Int}(u, I_1) + \text{Int}(u, I_2)$  and if  $I_1 \subseteq I_2$ , then  $\text{Int}(u, I_1 \setminus I_2) = \text{Int}(u, I_1) - \text{Int}(u, I_2)$ . As we will see, this linearity property permits us to develop efficient algorithms for computing interestingness for arbitrary intervals, after moderate amounts of preprocessing.

# Folksonomy з точки Ontology Learning, Formal Concept Analysis and its triadic version, and the mining of closed itemsets

## Discovering shared conceptualizations in folksonomies

Robert Jäschke<sup>a,b</sup>, Andreas Hotho<sup>a</sup>, Christoph Schmitz<sup>a</sup>,  
Bernhard Ganter<sup>c</sup>, Gerd Stumme<sup>a,b,\*</sup>

<sup>a</sup> Knowledge & Data Engineering Group, University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany<sup>†</sup>

<sup>b</sup> Research Center L3S, Appelstr. 9a, 30167 Hannover, Germany<sup>‡</sup>

<sup>c</sup> Institute for Algebra, Dresden University of Technology, 01062 Dresden, Germany<sup>‡</sup>

Received 8 June 2007; received in revised form 30 August 2007; accepted 6 November 2007

Available online 17 November 2007

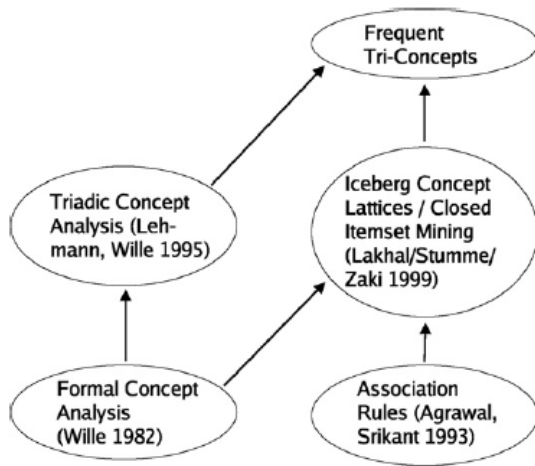


Fig. 1. History of iceberg tri-lattices.

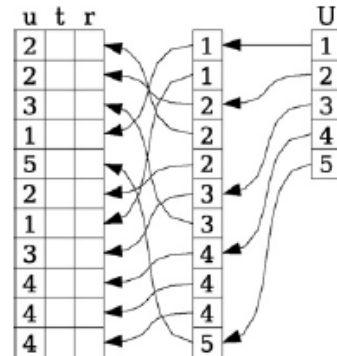


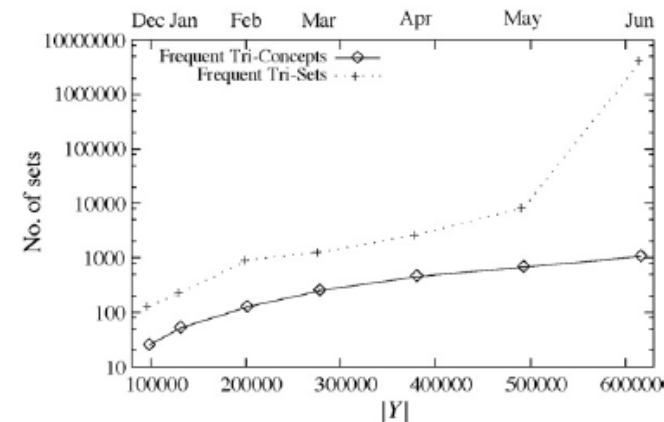
Fig. 3. Accessing triples in sorted order.

**Algorithm 1.** The TRIAS algorithm for mining all frequent tri-concepts

**TRIAS**( $U, T, R, Y, \tau_u, \tau_t, \tau_r$ )

1. **begin**
2.  $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$
3.  $(A, I) := \text{FirstFrequentConcept}((U, T \times R, \tilde{Y}), \tau_u)$
4. **repeat**
5.   **if**  $|I| \geq \tau_t \cdot \tau_r$  **then begin**
6.      $(B, C) := \text{FirstFrequentConcept}((T, R, I), \tau_t)$
7.     **repeat**
8.       **if**  $|C| \geq \tau_r$  **then**
9.         **if**  $A = (B \times C)^{\tilde{Y}}$  **then**  $\text{output}(A, B, C)$
10.       **until not**  $\text{NextFrequentConcept}((B, C), (T, R, I), \tau_t)$
11.     **endif**
12.   **until not**  $\text{NextFrequentConcept}((A, I), (U, T \times R, \tilde{Y}), \tau_u)$
13. **end**

R. Jäschke et al. / Web Semantics: Science, Services and Agents on the World Wide Web 6 (2008) 38–53



# Folksonomy з точки зору теорії ґраток та Formal Concept Analysis

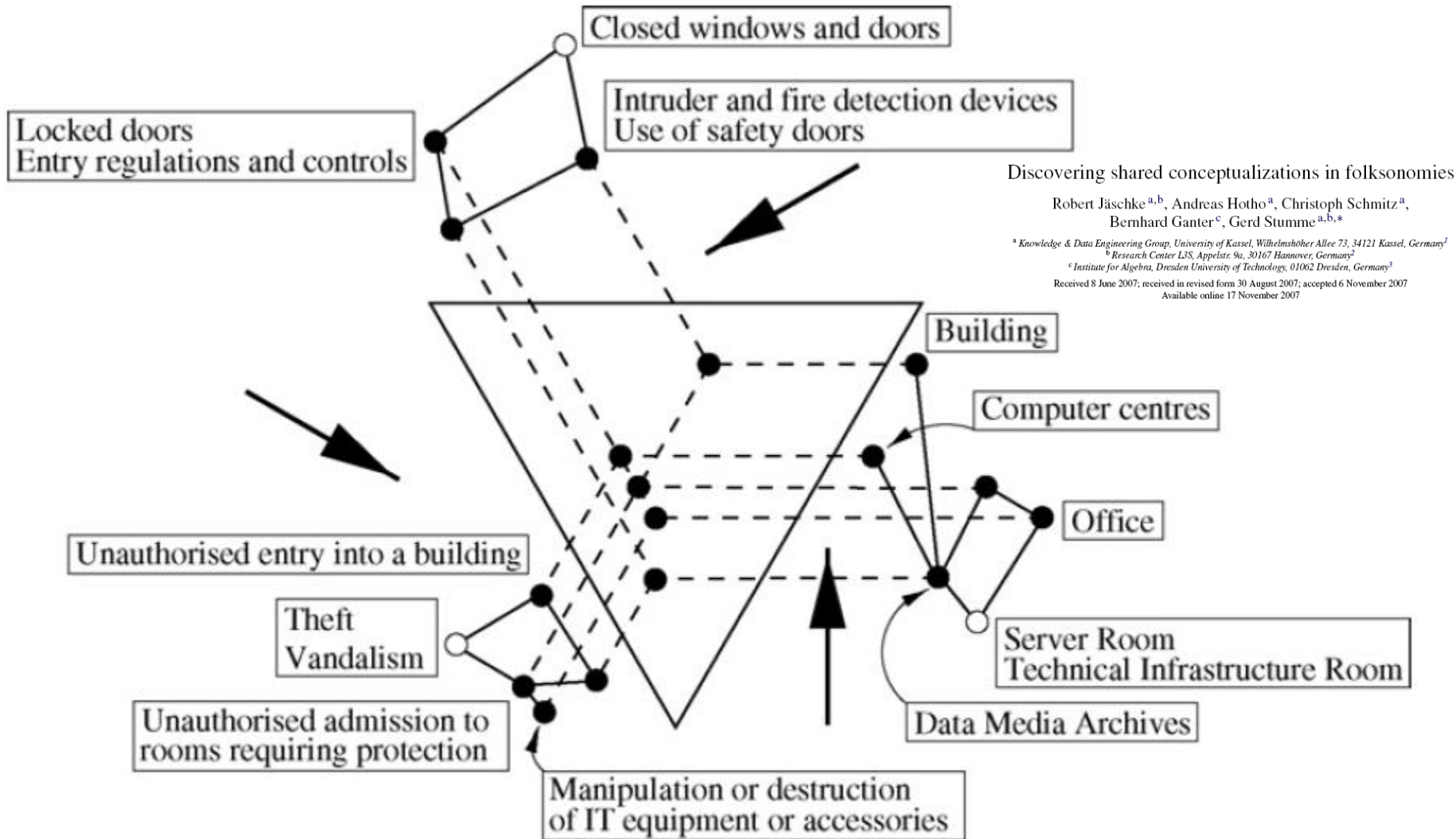
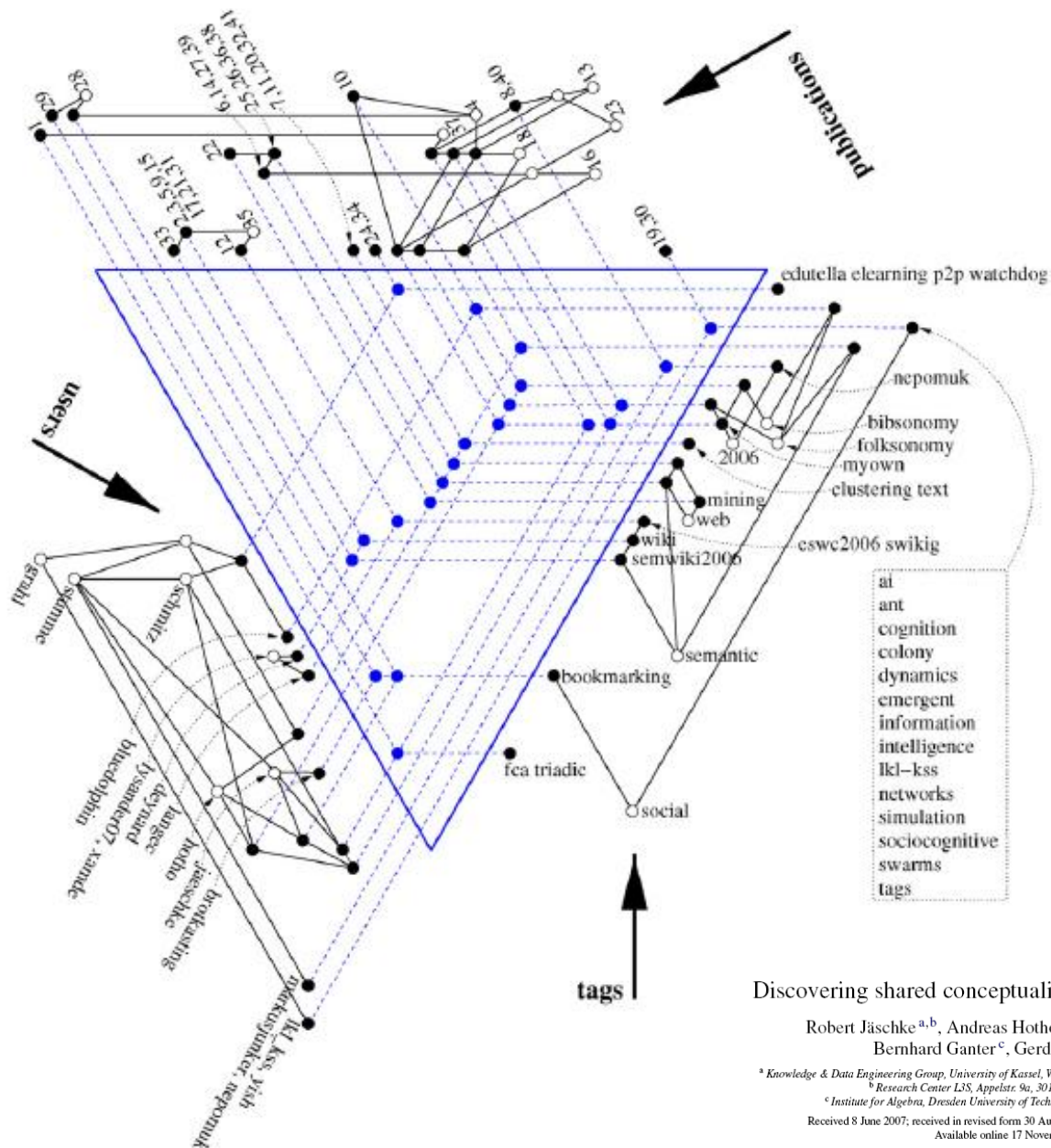


Fig. 7. All frequent tri-concepts of the IT Baseline Security Manual for  $\tau_u = \tau_t = \tau_r = 3$ .

# Folksonomy з точки зору теорії ґраток та Formal Concept Analysis



Discovering shared conceptualizations in folksonomies

Robert Jäschke<sup>a,b</sup>, Andreas Hotho<sup>a</sup>, Christoph Schmitz<sup>a</sup>,  
Bernhard Ganter<sup>c</sup>, Gerd Stumme<sup>a,b,\*</sup>

<sup>a</sup> Knowledge & Data Engineering Group, University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany<sup>1</sup>

<sup>b</sup> Research Center L3S, Appelstr. 9a, 30167 Hannover, Germany<sup>2</sup>

<sup>c</sup> Institute for Algebra, Dresden University of Technology, 01062 Dresden, Germany<sup>3</sup>

Received 8 June 2007; received in revised form 30 August 2007; accepted 6 November 2007

Available online 17 November 2007

Fig. 9. All frequent tri-concepts of the BibSonomy publications for  $\tau_u = 3$ ,  $\tau_t = 2$ ,  $\tau_r = 2$ .



# A Triadic Approach to Formal Concept Analysis by [Fritz Lehmann](#), [Rudolf Wille](#), 1995

---

## A Triadic Approach to Formal Concept Analysis

**Source** [Lecture Notes In Computer Science; Vol. 954](#) [archive](#)  
[Proceedings of the Third International Conference on Conceptual Structures: Applications, Implementation and Theory](#) [table of contents](#)  
Pages: 32 - 43  
Year of Publication: 1995  
ISBN:3-540-60161-9

**Authors** [Fritz Lehmann](#)  
[Rudolf Wille](#)

**Publisher** Springer-Verlag London, UK

**Bibliometrics** Downloads (6 Weeks): n/a, Downloads (12 Months): n/a, Citation Count: 5

---

**Additional Information:** [cited by](#) [collaborative colleagues](#)

**Tools and Actions:** [Review this Article](#)  
[Save this Article to a Binder](#) Display Formats: [BibTex](#) [EndNote](#) [ACM Ref](#)

---

### ↑ CITED BY 5

[Min Liu , Mingwen Shao , Wenxiu Zhang , Cheng Wu, Reduction method for concept lattices based on rough set theory and its application, Computers & Mathematics with Applications, v.53 n.9, p.1390-1410, May, 2007](#)

[Jitender S. Deogun , Jamil Saquer, Monotone concepts for formal concept analysis, Discrete Applied Mathematics, v.144 n.1-2, p.70-78, 30 November 2004](#)

[Ciro Cattuto aff3n2 , Christoph Schmitz , Andrea Baldassarri , Vito D. P. Servedio aff2n3 , Vittorio Loreto aff2n3 , Andreas Hotho , Miranda Grahl , Gerd Stumme, Network properties of folksonomies, AI Communications, v.20 n.4, p.245-262, December 2007](#)

[Robert Jäschke , Andreas Hotho , Christoph Schmitz , Bernhard Ganter , Gerd Stumme, Discovering shared conceptualizations in folksonomies, Web Semantics: Science, Services and Agents on the World Wide Web, v.6 n.1, p.38-53, February, 2008](#)

[Gerd Stumme, Off to new shores: conceptual knowledge discovery and processing, International Journal of Human-Computer Studies, v.59 n.3, p.287-325, September 2003](#)

### ↑ Collaborative Colleagues:

Fritz Lehmann: [colleagues](#)

Rudolf Wille: [colleagues](#)

# A Triadic Approach to Formal Concept Analysis by [Fritz Lehmann](#), [Rudolf Wille](#), 1995

	1					2					3					4					5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
3	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x			x	x			
4	x	x	x	x	x	x	x	x	x		x	x	x			x	x				x				
5	x	x	x	x		x	x	x			x	x				x									

Fig. 1. The triadic 5-chain context  $\mathbb{K}_5^c$

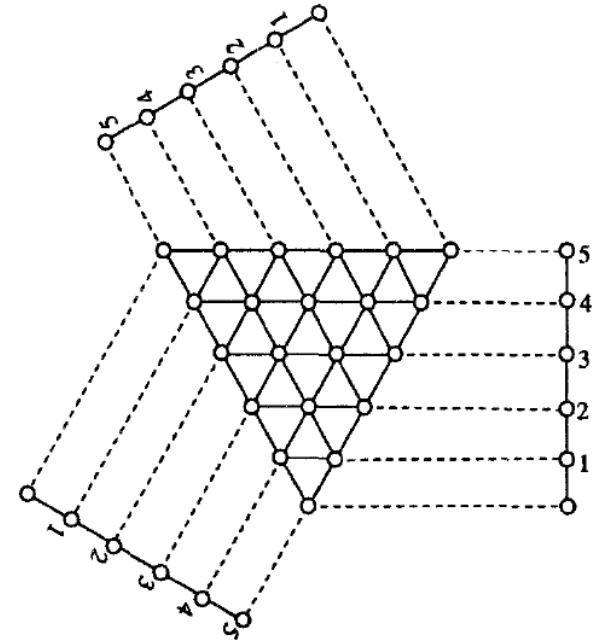
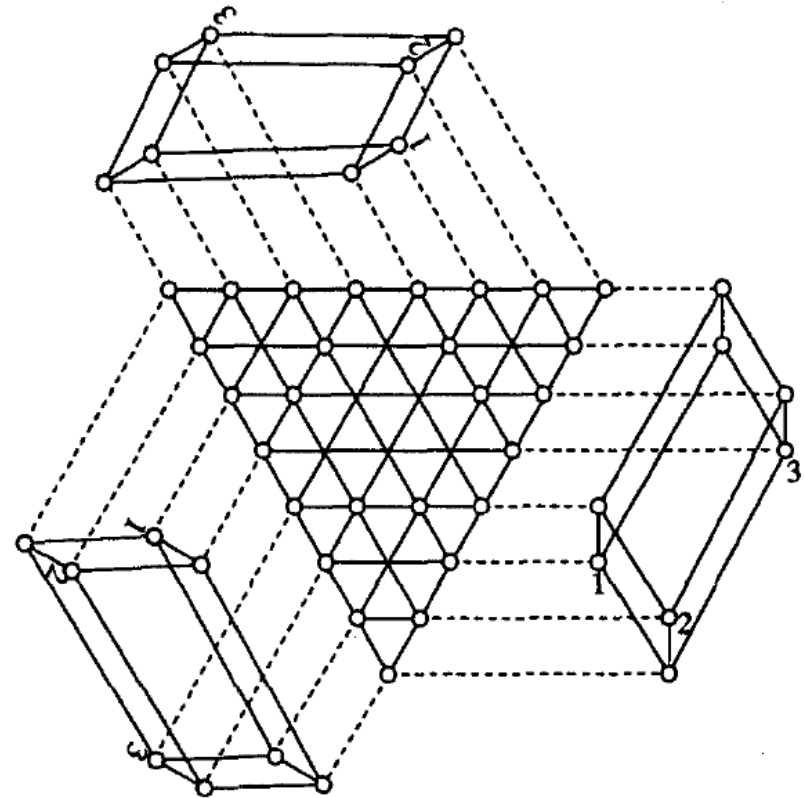


Fig. 2. A triadic diagram of the triadic 5-chain  $\underline{\mathbb{T}}(\mathbb{K}_5^c)$

# A Triadic Approach to Formal Concept Analysis by [Fritz Lehmann](#), [Rudolf Wille](#), 1995

	1			2			3		
	1	2	3	1	2	3	1	2	3
1		x	x	x	x	x	x	x	x
2	x	x	x	x			x	x	x
3	x	x	x	x	x	x	x	x	

**Fig. 3.** The triadic power set context  $\mathbb{K}_{\{1,2,3\}}^b$



**Fig. 4.** The power set trilattice  $\mathfrak{T}(\mathbb{K}_{\{1,2,3\}}^b)$

# A Triadic Approach to Formal Concept Analysis by [Fritz Lehmann, Rudolf Wille](#), 1995

	a						b						c						d					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
A	x	x					x	x					x	x					x	x	x			
B	x												x	x					x	x				x
C	x	x	x	x	x	x	x	x	x	x	x	x	x	x					x	x	x			x

- A: Superordinate stereotypes
  - B: Unspecified
  - C: Subtypes
  - a: Recall
  - b: Impression Formation
  - c: Behavior Prediction
  - d: Evaluation
- 1. Physical appearance
  - 2. Political beliefs
  - 3. Attitudes
  - 4. Behavior
  - 5. Traits
  - 6. Situations

Fig. 5. A triadic context of experimental data

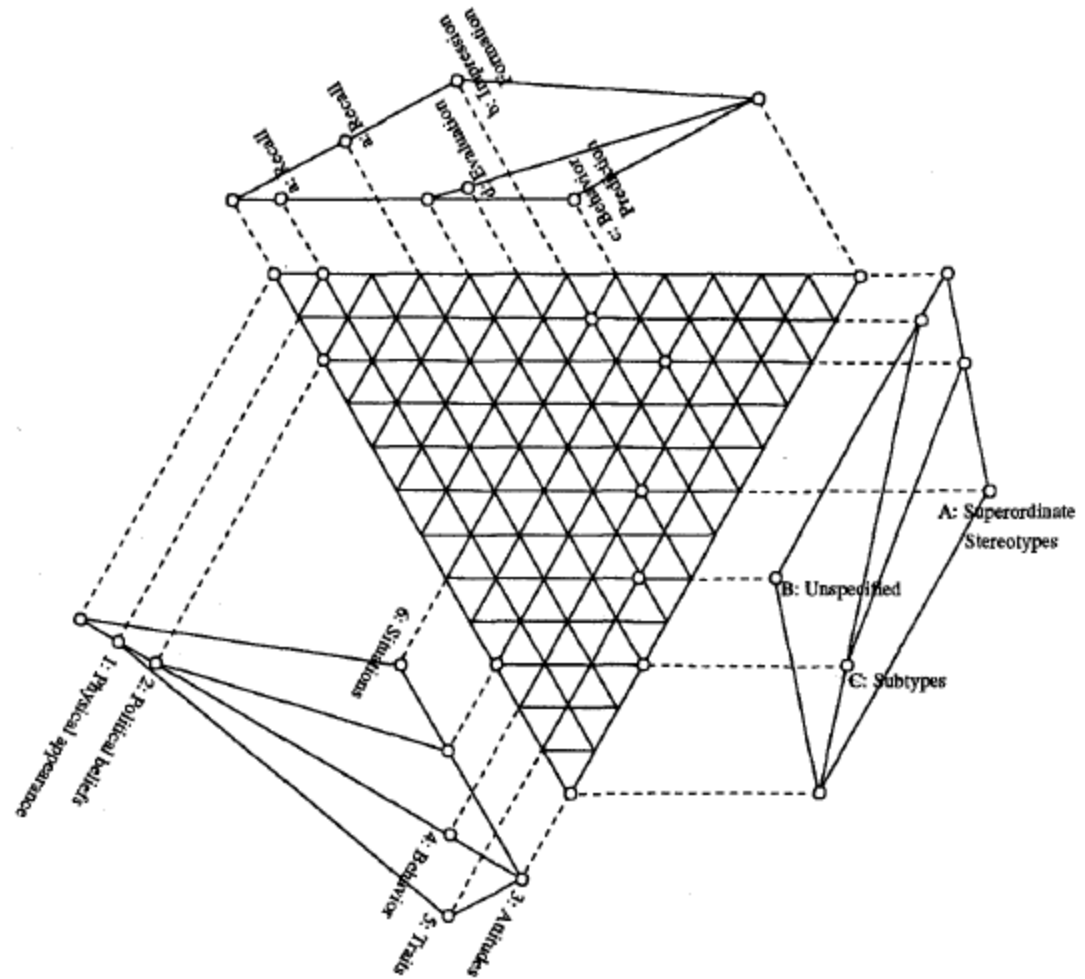


Fig. 6. A triadic diagram of the 'triadic' relational structure of the context in Figure 5

# Bibsonomy – погляд з середини

## Semantics of the BibSonomy URL scheme

Since group visibility rights (see Section 4.4) make the explanation much more complicated, they are mostly disregarded in this section, as well as in the formal model.

All URLs described here are relative to `http://www.bibsonomy.org`. There are system pages like `/help`, `/settings` or `/post_bookmark` which are necessary for the usage of BibSonomy, but their semantic is straightforward hence they are omitted here. The following list describes the contents  $C$  of all pages which show posts in BibSonomy:

`/tag/t1...tn` Shows every post which has all of the tags  $t_1, \dots, t_n$  attached:

$$C_{t_1, \dots, t_n} := \{(u, S, r) \in P \mid \{t_1, \dots, t_n\} \subseteq S\} \quad (1)$$

`/user/u` Shows all posts of user  $u$ :

$$C_u := \{(\hat{u}, S, r) \in P \mid \hat{u} = u\} \quad (2)$$

`/user/u/t1...tn` Shows every post of user  $u$  which has all of the tags  $t_1, \dots, t_n$  attached:

$$C_{u, t_1, \dots, t_n} := \{(\hat{u}, S, r) \in P \mid \hat{u} = u, \{t_1, \dots, t_n\} \subseteq S\} \quad (3)$$

`/concept/user/u/t1...tn` Shows every post of user  $u$  which has for every tag  $t \in \{t_1, \dots, t_n\}$  at least one of its subtags or  $t$  itself attached (see also Section 4.4):

$$C_{u, t_1, \dots, t_n} := \{(\hat{u}, S, r) \in P \mid \hat{u} = u, \forall t_i (i = 1, \dots, n) \exists t \in S : (\hat{u}, t, t_i) \in \prec \vee t = t_i\} \quad (4)$$

# Bibsonomy – погляд з середини

**/url/r** If  $r$  is a bookmark: Shows all posts of the resource  $r$ :

$$C_r := \{(u, S, \hat{r}) \in P \mid \hat{r} = r\} \quad (5)$$

**/url/r/u** If  $r$  is a bookmark: Shows the post of user  $u$  of the resource  $r$ :

$$C_{r,u} := \{(\hat{u}, S, \hat{r}) \in P \mid \hat{r} = r, \hat{u} = u\} \quad (6)$$

**/bibtex/r** If  $r$  is a literature reference: Shows all posts of the resource  $r$ :

$$C_r := \{(u, S, \hat{r}) \in P \mid \hat{r} = r\} \quad (7)$$

**/bibtex/r/u** If  $r$  is a literature reference: Shows the post of user  $u$  of the resource  $r$ :

$$C_{r,u} := \{(\hat{u}, S, \hat{r}) \in P \mid \hat{r} = r, \hat{u} = u\} \quad (8)$$

**/group** Shows all groups of the system. More on groups can be found in Section 4.4.

**/group/g** Shows all posts of all users belonging to the group  $g$ :

$$C_g := \{(u, S, r) \in P \mid u \in g\} \quad (9)$$

**/group/g/t<sub>1</sub>...t<sub>n</sub>** Shows every post which has all of the tags  $t_1, \dots, t_n$  attached and where the user belongs to group  $g$ :

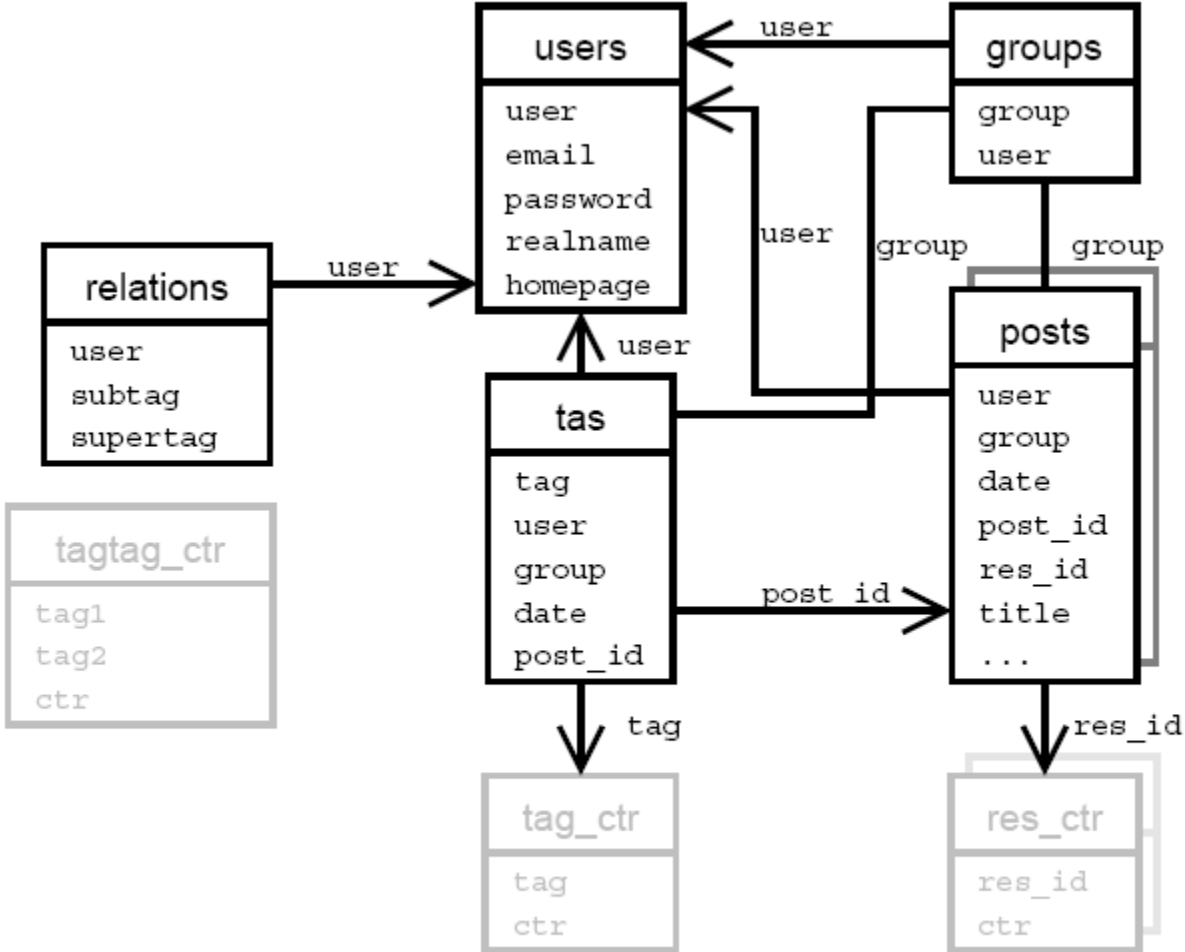
$$C_{g,t_1,\dots,t_n} := \{(u, S, r) \in P \mid u \in g, \{t_1, \dots, t_n\} \subseteq S\} \quad (10)$$

# Bibsonomy – погляд з середини

- `/viewable/g` Shows all posts which are set viewable for members of the group  $g$ .
- `/viewable/g/t1...tn` Shows all posts which are set viewable for members of the group  $g$  and which have all of the tags  $t_1, \dots, t_n$  attached.
- `/search/s` Shows all resources, whose full text matches the search expression  $s$ , which is interpreted by the MySQL full text search capability.<sup>11</sup>
- `/download` Shows all publication posts which the user has picked in the shopping basket, as described in section 4.4.
- `/popular` Shows the 100 resources posted most often amongst the last 1000 posts. (Note that these numbers are subject to change.)
- `/` This is the home page of BibSonomy, it shows the entries posted most recently.

An interesting feature, described in section 4.4, is that all paths of URLs described above, can be prepended by a string which changes the output format. In general, posts are shown as HTML lists surrounded by navigation elements and a tag cloud (as seen in Figure 1), but this feature allows the user to get her output in formats like  $\text{BIB}_{\text{T}}\text{E}_\text{X}$  or as an RSS feed.

# Bibsonomy – погляд з середини



Relational schema of the most important tables.



# Математичні основи терміна Folksonomy (2.1)

**Взаємна (обопільна, спільна) контекстуалізація гіперграфу** – як елементи даних множин набувають своєї власної семантики через зв'язки між ними.



## Mutual contextualization in tripartite graphs of folksonomies



Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt  
{cma06r, nmg, nrs}@ecs.soton.ac.uk

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

### Overview

#### Objectives

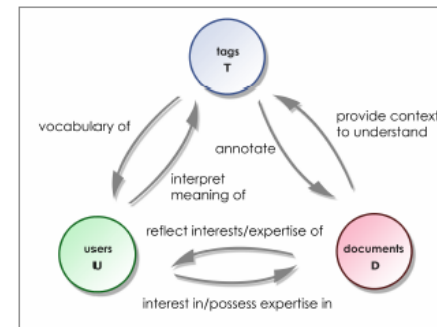
- To understand the dynamics in folksonomies
- To investigate how semantics can be extracted from folksonomies
- To study the potential applications of the information extracted from folksonomies

#### Methods

- Apply network analysis techniques on graphs of folksonomy
- Apply results on various applications: e.g. tag disambiguation and recommender systems

### Mutual Contextualization

- What does a user ID, a tag or a URL represents?
- Folksonomies contain interrelations between these entities.
- The entities provide a context for each other so that their semantics can be understood.
- Hence, we can acquire a better understanding of the entities by exploring the associations between them.



A folksonomy  $F = (U, T, D, A)$

$A$  is a set of annotations:  $A \subseteq U \times T \times D$

# Математичні основи терміна Folksonomy (2.2)

## Study Individual Elements in a Folksonomy

- To study individual elements in a folksonomy, we extract bipartite graphs from the folksonomy hypergraph.
- For example, for a particular user, we can extract all the tags used and all the documents tagged by this user; this is similar for a tag or a document.
- We can then fold the bipartite graphs into different one-mode networks for analysis.
- Findings are expected to be useful and beneficial to various applications on the Web.



**A user ( $u$ )**

Bipartite graph  $TD_u$

$$TD_u = \langle T \cup D, E_{TD} \rangle$$

$$E_{TD} = \{ \{t, d\} \mid \{u, t, d\} \in A \}$$



**A tag ( $t$ )**

Bipartite graph  $UD_t$

$$UD_t = \langle U \cup D, E_{UD} \rangle$$

$$E_{UD} = \{ \{u, d\} \mid \{u, t, d\} \in A \}$$



**A document ( $d$ )**

Bipartite graph  $UT_d$

$$UT_d = \langle U \cup T, E_{UT} \rangle$$

$$E_{UT} = \{ \{u, t\} \mid \{u, t, d\} \in A \}$$

Folded into **one-mode networks** by **adjacency matrix multiplication**

- tag network
- document network

- user network
- document network

- tag network
- user network

### Potential Applications

- Disambiguate tags with multiple meanings
- Constructing of user profiles
- Generating of personal ontologies
- Improving recommender systems
- Deriving social networks of shared interests
- Generating annotations for documents

# Математичні основи терміна Folksonomy (2.3)

## Folksonomy (A hypergraph)

$$F = \langle U, T, D, A \rangle; A \subseteq U \times T \times D$$



A Tag

Bipartite graph  $UD_t$

$$UD_t = \langle U \cup D, E_{UD} \rangle$$

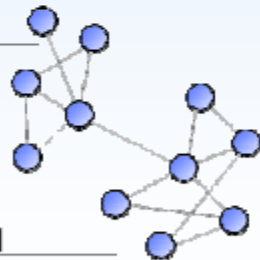
$$E_{UD} = \{ \{u, d\} \mid \{u, t, d\} \in A \}$$

adjacency matrix multiplication

user

edge weight =  
# of documents tagged

A weighted network of users



edge weight =  
# of users tagged the documents

documents

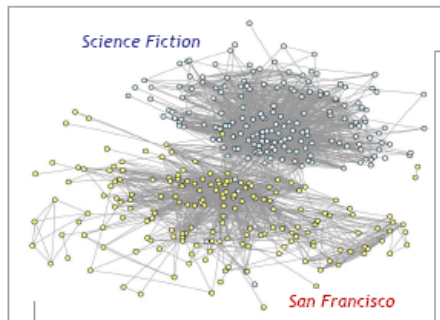
A weighted network of documents



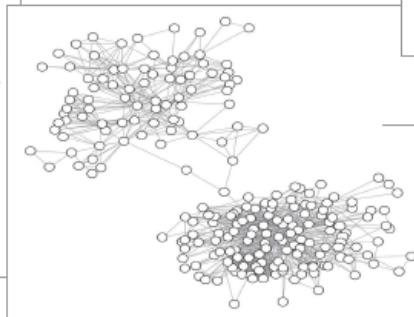
# Математичні основи терміна Folksonomy (2.4)

## Case Study: Tag Disambiguation

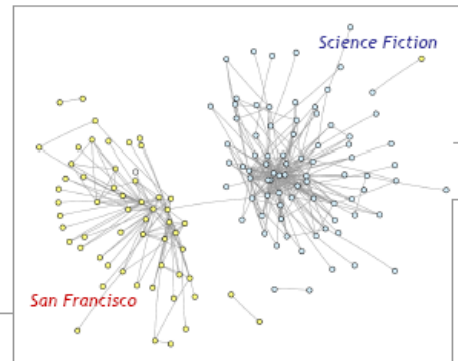
- To investigate tag semantics, we try to examine the document and user networks of a single tag.
- Focusing on the tag *sf*, we extract the document network and user network.
- We study the possibility of automatic tag disambiguation by examining the topology of the networks.
- Results show that majority of the users use the tag to refer to one meaning only, resulting in different clusters of documents.



Network of documents being tagged by the tag *sf*



Network of users who have used the tag *sf*



Network of most frequently used tags



Network of documents, edges with weights < 2 removed

A paper on this work will be presented in the [International Workshop of Emergent Semantics and Ontology Evolution \(ESOE 2007\)](#) on 12 Nov. You can also visit <http://www.ecs.soton.ac.uk/~cmay06r/> for more information.

Doctoral Consortium, International Semantic Web Conference 2007, Busan, South Korea, 11-15 Nov 2007

**Prof Nigel R Shadbolt**  
School of Electronics and Computer Science  
University of Southampton  
Southampton  
SO17 1BJ  
United Kingdom



**Dr Nicholas Gibbins**  
School of Electronics and Computer Science  
University of Southampton  
Southampton  
SO17 1BJ  
United Kingdom



# Програми для візуалізації тегів (TouchGraph) – tag: “Folksonomy” на Google.com

TouchGraph Google Browser - Mozilla Firefox

http://www.touchgraph.com/TGGoogleBrowser.php?start="web 2.0"&signed=false

Info Help

Web 2.0 - Wikipedia, the free encyclopedia

[http://en.wikipedia.org/wiki/Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0)

Web 2.0 is a term describing changing trends in the use of World Wide Web technology and web design that aims to enhance creativity, secure information ...

Filter

Show Hidden  Name

Name	URL	Sim#
Web 2.0 Expo 2007, April ...	conferences...	1
Web 2.0 Expo April 17, 20...	blog.damonl...	1
Flickr: "web2expo"	flickr.com/ph...	1
Web 2.0 Expo 2007, April ...	conferences...	1
web2expo: Blogs, Photos...	technorati.co...	1
Web 2.0 Expo   San Franc...	newsletter.w...	1
Web 2.0 Workgroup - A n...	web2workg...	10
Refresh Orlando	refreshorlan...	1
Scobleizer -- Tech geek b...	scobleizer.c...	1
Flickr: "web20logo"	flickr.com/ph...	1
LinkedIn: Groups: Sign In	https://www.l...	1
XING - Groups - All Group...	xing.com/net...	1
Web 2.0 Workgroup - Bok...	bokardo.co...	1
http://www.pctechnicians...	pctechnician...	1
Web 2.0 Workgroup Upd...	readrwrite...	1
The Web 2.0 WorkGroup	techcrunch.c...	1
Web 2.0 WorkGroup Now...	techcrunch.c...	1
Web 2.0 Summit 2008 - ...	en.oreilly.co...	0
Web 2.0 Expo San Franci...	en.oreilly.co...	0
Web 2.0 - Wikipedia, the f...	en.wikipedia...	10

Keywords or URL "web 2.0" Go Edit Keep Hide Expand

Label Shows  Title  URL Text Over Icon Tools

Zoom: Spacing:

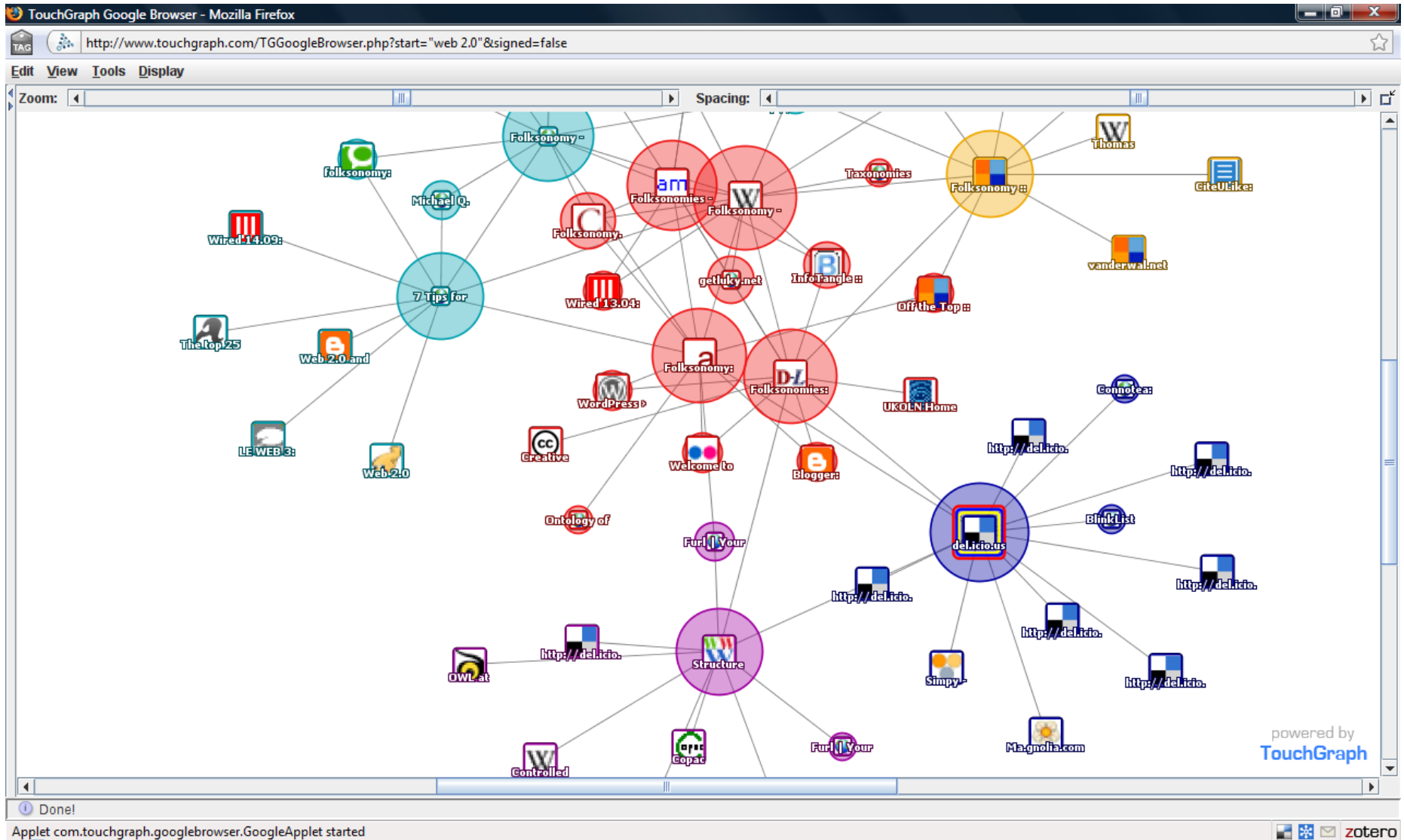
powered by TouchGraph

Done!

Applet com.touchgraph.googlebrowser.GoogleApplet started

zotero

# Програми для візуалізації тегів (TouchGraph) – tag: “Folksonomy” на Google.com



# Програми для візуалізації тегів (TouchGraph) – tag: “Formal Concept Analysis” на amazon.com (книги)

TouchGraph Amazon Browser - Mozilla Firefox

http://www.touchgraph.com/TGAmazonBrowser.php?query\_type=Book&keywords=FCA

File Edit View Tools Display

Info Help

Title Formal Concept Analysis: Foundations and Applications (Lecture Notes in Computer Science)

Sale Price \$68.03

List Price \$84.95

Author --

Publisher Springer

[View/Buy on Amazon](#)

Search Books Formal Concept Analysis Go Edit Keep Hide Expand

Display Book Author Subject Publisher Links Tools

Zoom: Spacing:

Book Author Subject Publisher Queries

Filter

Show Hidden Search Names

X	Title	Sim#	%Rnk	Date	Price
X	Formal Concept Analysis	10	94	2005	\$68
X	Programming Collective Intelligence	4	57	2007	\$26
X	Pattern Recognition and Machine Learning	3	71	2007	\$67
X	Knowledge Representation and Reasoning	3	84	2004	\$64
X	Structure and Interpretation of Computer Science	2	90	1996	\$59
X	Introduction to Lattices and Order	2	93	2002	\$43
X	Concept Data Analysis	2	95	2004	\$111
X	Social Network Analysis	1	74	1994	\$42
X	Concept Lattices	1	95	2004	\$64
X	Conceptual Structures	1	95	2000	\$79
X	Formal Concept Analysis	1	96	2006	\$53

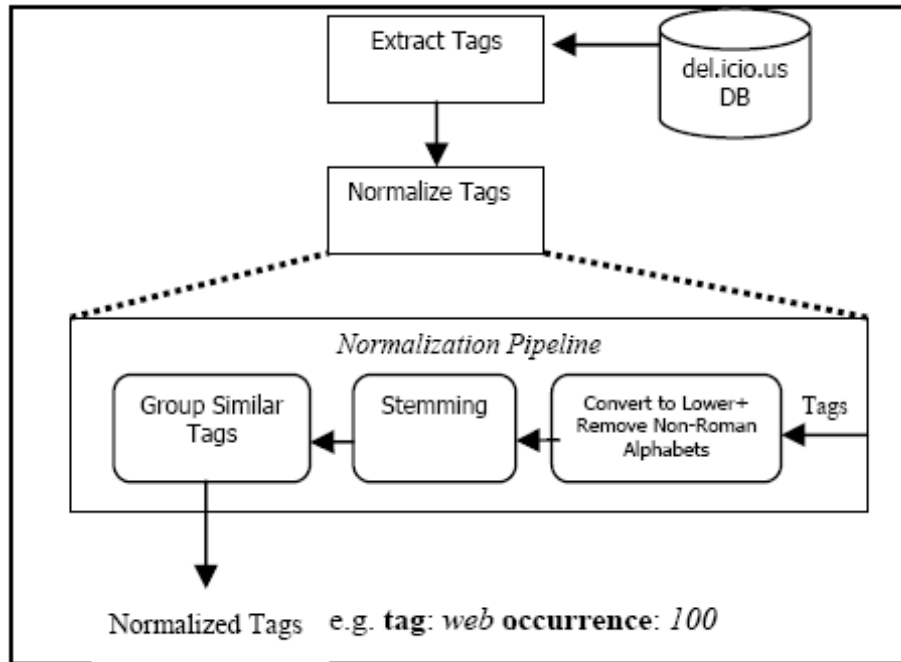
powered by TouchGraph

Done!

Applet com.touchgraph.amazonbrowser.AmazonApplet started

zotero

# Метод аналізу фолксономічних шаблонів (Folksonomic Patterns)



*Al-Khalifa, H. S. and Davis, H. C. (2007) Towards Better Understanding of Folksonomic Patterns. In: Eighteenth ACM Conference on Hypertext and Hypermedia (HT 07), 10-12 September, Manchester, UK.*

Classification	P	S	F
abbreviations & acronyms	10%	-	1%
complete words	27%	86%	96%
compound words	57%	13%	3%
typos & non-English words	6%	1%	-

The normalization pipeline contains four filters:

1. **Lower-case filter:** Tags are converted to lower case so that string manipulation (e.g. comparison) can be applied easily,
2. **Non-English filter:** Non-Roman alphabets are dropped; this step is to insure that only English-like tags are present when doing the analysis,
3. **Stemming filter:** Tags are stemmed (e.g. plurals converted to singular) using a modified version of the Porter Stemmer.
4. **Grouping Similar Tags filter:** identical tags and substrings are grouped and their occurrence is counted.



# Метод аналізу фолксономічних шаблонів (Folksonomic Patterns)

*Al-Khalifa, H. S. and Davis, H. C. (2007) Towards Better Understanding of Folksonomic Patterns. In: Eighteenth ACM Conference on Hypertext and Hypermedia (HT 07), 10-12 September, Manchester, UK.*

## Data Set

Our data set consist of a sample of 100 randomly selected bookmarked web resources from the domain of CSS with a total number of 72,458 posts (i.e. a post is a resource tagged by a user). The total number of tags before normalization was 245,892, and the number of tags after normalization was 10,900. These 10,900 folksonomy tags were manually inspected and classified based on the Sen et al. [1] classification, as will be described in the next section.

*[1] Sen, S., S. Lam, A. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, M. Harper, and J. Riedl. Tagging, communities, vocabulary, evolution. In Proceedings of CSCW '06. 2006. ACM.*

## **(P)ersonal tags**

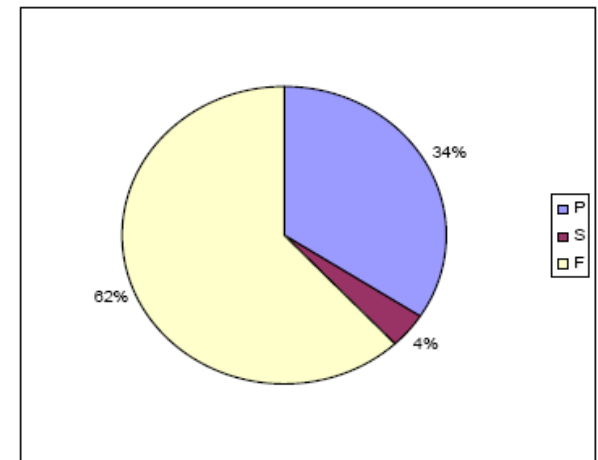
Self-reference tags classification, includes any tag that has to do with the user own interest. Such as dates e.g. (January, monthly and night), names e.g (tojack) and own reference e.g. (mylink, mysite and myblog). These tags usually appear once or twice among all tags in a given bookmarked web resource.

## **(S)ubjective tags**

These are tags that express people's opinions on the bookmarked web resource. Although these tags constitute a small portion of the inspected folksonomy tags (i.e. 4%), an in depth inspection was carried out to analyze them.

## **(F)actual tags**

These are tags which identify 'facts' about the described web resource such as people, places, or concepts. A more specific rough classification can be: *web resource title/URL/author, synonyms (either near or far), rights/language, compound tags, generic, acronyms, spelling variation and other areas of application or usage.*



# Метод аналізу фолксономічних шаблонів (Folksonomic Patterns)

*Al-Khalifa, H. S. and Davis, H. C. (2007) Towards Better Understanding of Folksonomic Patterns. In: Eighteenth ACM Conference on Hypertext and Hypermedia (HT 07), 10-12 September, Manchester, UK.*

## CONCLUSION AND FUTURE WORK

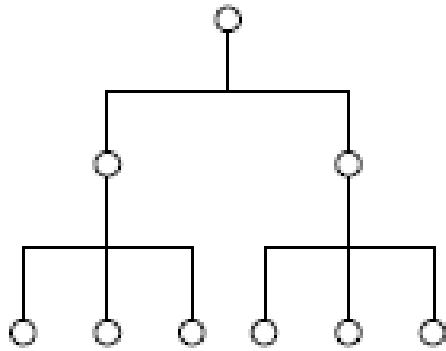
In previous work [1] we have compared folksonomy tags to keywords extracted using context based keyword extraction technique, and demonstrated the improved value of the folksonomy tags over context-based keywords. However, in this work we showed that folksonomy tags have the potential to be transformed into meaningful metadata by classifying them into semantic categories. This was proven by showing that a great shear of folksonomy tags were indeed factual. This successful classification of folksonomy tags into meaningful semantics can open the doors for future research in processing and using folksonomy tags in creating structured metadata that adhere to pre-defined ontologies, as we have done in [6]. Finally, to the best of our knowledge this is the first time an experiment with this magnitude was conducted manually to inspect the generic semantics of folksonomy tags (cf. [3]).

*[3] Kipp, M.E. Exploring the context of user, creator and intermediate tagging. In IA Summit 2006. 2006. Canada.*



**Dr Hugh C Davis**  
School of Electronics and Computer Science  
University of Southampton  
Southampton  
SO17 1BJ  
United Kingdom

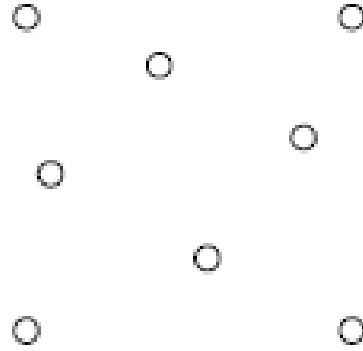
# Три основні підходи до класифікації (індексації) даних



Classification

Тахоному (класифікація)  
Чітко виражена ієрархічна структура

- Dewey Decimal Classification (DDC)
- Класифікація бібліотеки конгресу США
- Радянська класифікація



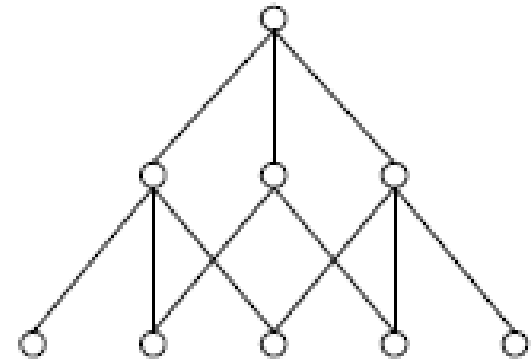
Unconnected Tags

Tagging systems (**Folksonomy**)

“Плоска” структура тегів не має зв’язків

Власне вона і є основним недоліком “народної класифікації”

Див. статтю: “Collaborative thesaurus tagging the Wikipedia way”  
Jakob Voss, Wikimedia Deutschland e.V. (ver.2) - 2006-04-27



Thesaurus

Технологія тлумачного словника (thesaurus) – відома більше, ніж 50 років.  
Новий підхід в Web 2.0 – Вікіпедія (wikipedia)

Зберігаючи чітку ієрархічну структуру, елементи додатково мають “міжрівневі” зв’язки

# Висновки

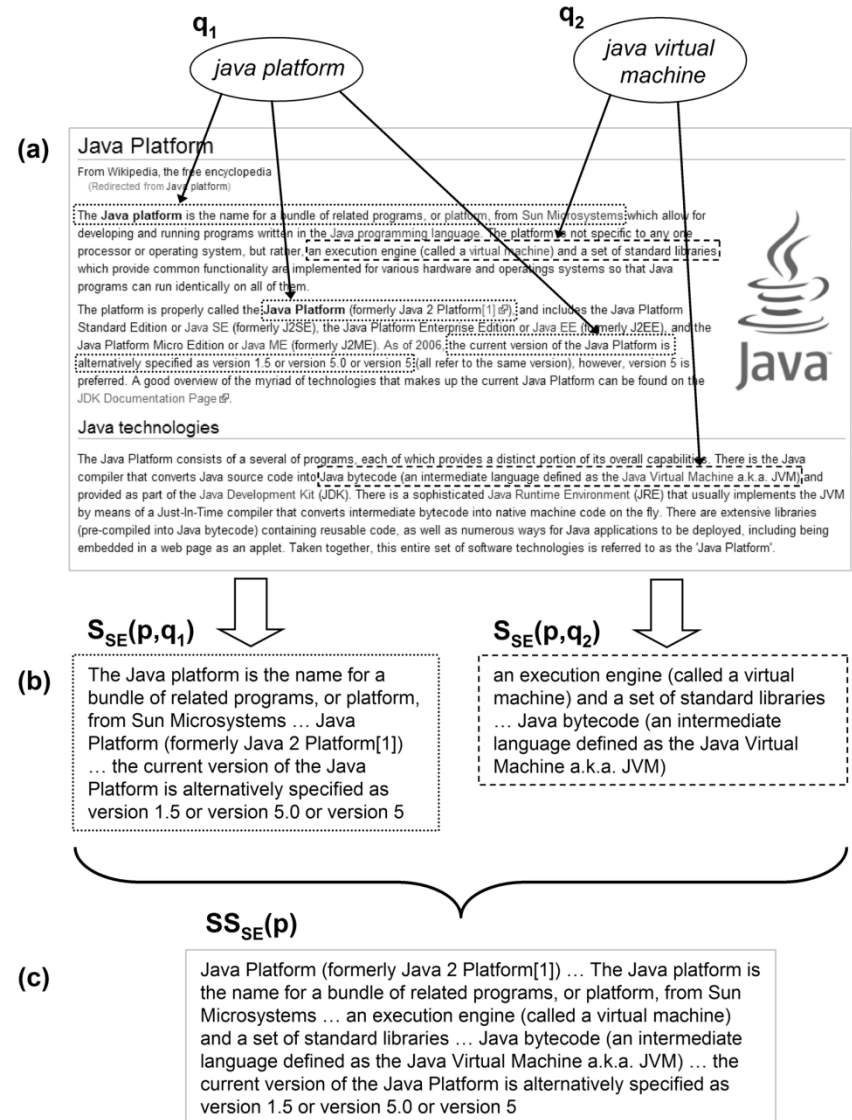
- Проблема синонімів, ононімів та інше.
- Різний рівень словникового запасу користувача та його ставлення до створення тегів того чи іншого ресурсу. Створення ефективних алгоритмів фільтрування від спаму.
- Удосконалити “плоску” структуру тегів шляхом класифікації слів та зв'язків між ними.
- Семантичне “збагачення” простору тегів “народної класифікації”. Розробка ефективних алгоритмів.
- Виявлення прихованих ком'юніті за інтересами та можливість співпраці між ними.
- Створення ефективних алгоритмів пошуку та ранжування на основі даних “народної класифікації”.

# Погляд у майбутнє, тенденції, перспективи – Social Summarization

## Від соціальних закладок (Social Bookmarks) до соціальних анотацій (Social Summarization) - ресурсів в інтернеті

На рисунку показано соціальну анотацію.

- (a) Сторінка  $p$  асоціюється з множиною  $a$  запитів, що використовує доступ до неї (чи використанні **теги** до закладок  $p$ ), в цьому випадку  $q_1$  і  $q_2$ .
- (b) Для кожного запиту пошуковик буде видавати **a query-focused snippet**,  $S_{SE}(p, q_1)$  and  $S_{SE}(p, q_2)$ , що складається з чутливих до контенту фрагментів сторінки, що пов'язаний з пошуком.
- (c) Fragments are **scored** and rank-ordered to produce the final social summary,  $SS_{SE}(p)$



"From Social Bookmarking to Social Summarization: An Experiment in Community-Based Summary Generation" Oisin Boydell, Barry Smyth (Adaptive Information Cluster School of Computer Science and Informatics University College Dublin), IUI'07, January 28–31, 2007, Honolulu, Hawaii, USA

Соціальні закладки,  
“народна класифікація” – складні мережі  
Social Bookmarks, **Folksonomies** – Complex Networks



- Дискусійна група розсилки  
<http://groups.google.com/group/ir-ua>  
Приєднуйтеся !

**Дякую за увагу**

2008 (с) Назаревич О.Б.  
Асистент каф. ВІ та КН  
taltek.te@gmail.com