TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

─────── \* ───────



TIỂU LUẬN MÔN HỌC

**Web 3.0 và các dịch vụ Internet thế hệ mới**

Đề tài: Chapter 12 - Social Bookmarking: An Overview of Folksonomies

Học viên thực hiện: **Bùi Nguyên Tùng** - CA170262 - TTM17A

**Phạm Duy Hưng** - CA170259 - TTM17A

**Lê Đức Dương**  - CB160559 - TTM16B

Giáo viên hướng dẫn: **TS. Phạm Huy Hoàng**

HÀ NỘI 1-2018

**MỤC LỤC**

[ĐẶT VẤN ĐỀ 3](#_Toc503993297)

[CHƯƠNG 1: KHÁI NIỆM FOLKSONOMY 4](#_Toc503993298)

[1. Social bookmarking là gì? 4](#_Toc503993299)

[2. Folksonomy là gì? 5](#_Toc503993300)

[3. Các thuộc tính quan trọng của folksonomy 6](#_Toc503993301)

[CHƯƠNG 2: THỰC TRẠNG, LỢI THẾ, VÀ THÁCH THỨC ĐỐI VỚI FOLKSONOMY 8](#_Toc503993302)

[1. Các dịch vụ social bookmarking nổi tiếng 8](#_Toc503993303)

[2. Lợi thế của Folksonomy 10](#_Toc503993304)

[3. Thách thức đối với Folksonomy 12](#_Toc503993305)

[CHƯƠNG 3: NGHIÊN CỨU GIẢI QUYẾT CÁC THÁCH THỨC 14](#_Toc503993306)

[1. Các giải pháp hiện nay 14](#_Toc503993307)

[2. Các hướng nghiên cứu phát triển folksonomy 16](#_Toc503993308)

[3. Kết luận 18](#_Toc503993309)

# ĐẶT VẤN ĐỀ

Kể từ khi xuất hiện Web 2.0 đã mở ra một thời kỳ mới cho các ứng dụng Web, trong đó Web được xem như là nền tảng (platform) trong việc phát triển ứng dụng, chia sẻ thông tin qua các dịch vụ cung cấp. Nếu như các ứng dụng thuộc thế hệ 1.0 chủ yếu là các website đóng của các hãng thông tấn hay các công ty nhằm cung cấp thông tin và tiếp cận với khách hàng một cách hiệu quả thì các ứng dụng web 2.0 đã phát triển dần sang một xu hướng khác mà ở đó yếu tố tham gia của người dùng đóng vai trò chủ đạo.

Các ứng dụng Web 2.0 đánh dấu những thay đổi quan trọng không chỉ ở sự đột phá trong nền tảng công nghệ mà còn cả ở cách thức sử dụng Internet của con người: mọi người cùng tham gia đóng góp cho xã hội “ảo” chứ không chỉ “duyệt và xem”. Sự thay đổi này vô hình chung đã hình thành nên những môi trường chia sẻ, công tác mà ở đó các cá nhân được liên kết chặt chẽ với nhau. Web 2.0 cũng cho ra đời nhiều khái niệm mới, trong đó có social bookmarking.

Trong một hệ thống Social bookmarking, người sử dụng sẽ lưu các đường dẫn tới các website mà họ muốn lưu trữ và chia sẻ cho cộng đồng, bạn bè bằng các từ khóa (tag) cá nhân. Các dịch vụ bookmarking hầu hết đều miễn phí, một số còn cung cấp cho người dùng một dung lượng bộ nhớ nhất định (ví dụ Flickr cho phép lưu trữ hình ảnh). Nhờ sự tiến bộ của các thuật toán và ứng dụng, các trang web cá nhân và các file thông tin (như hình ảnh, nhạc, video) đang được sắp xếp và chia sẻ một cách rộng rãi hơn tới cộng đồng.

Một biến thể phổ biến của social bookmarking là folksonomy. Folksonomy bao gồm sự cho phép người dùng thoải mái lựa chọn các từ khóa cá nhân và phân loại chúng nhằm mục đích dễ dàng truy cập lại trong tương lai và chia sẻ nội dung trên mạng. Các nội dung này khi được công khai sẽ được lên kết với các nội dung tương tự khác có cùng từ khóa, từ đó xây dựng nên một cộng đồng chung sở thích. Sự mở rộng và sẻ chia này đã làm nổi bật mặt xã hội của bookmarking.

Nội dung chương 12 mô tả rõ khái niệm của folksonomy, tác dụng của chúng, và phân tích những ưu điểm và thách thức đối với folksonomy. Làm thế nào tận dụng folksonomy nhằm đem lại một hiệu quả lợi nhuận tối đa trong xã hội mà vấn đề kết nối và quảng cáo đang đem lại một nguồn thu không nhỏ như hiện nay.

# CHƯƠNG 1: KHÁI NIỆM FOLKSONOMY

## Social bookmarking là gì?

Trước khi tìm hiểu khái niệm folksonomy, chúng ta cần hiểu social bookmarking là gì. Social bookmarking theo nghĩa đơn giản là đánh dấu trang cộng đồng giúp những người dùng sử dụng Internet để lưu trữ, quản lý, tìm kiếm địa chỉ trên những trang Web yêu thích trực tuyến trên Internet và chia sẻ nó với bạn bè dựa trên các công cụ trình duyệt Web thông qua các từ khóa phân loại. Những người được phép có thể xem tổng hợp một danh sách các link mà đã được phân loại theo các từ khóa (tag) hoặc thông qua sự trợ giúp của các công cụ tìm kiếm.

Dịch vụ social bookmarking (nghĩa là đánh dấu trang web xã hội) là một dịch vụ trực tuyến cho phép người dùng đăng, chú thích, chỉnh sửa và chia sẻ các đường dẫn đến các trang mạng với sự trợ giúp của các từ khóa phân loại. Dán nhãn từ khóa (tagging) là một trong những tính năng chính của các hệ thống social bookmarking, giúp người dùng có thể sắp xếp các đánh dấu liên kết của họ và mở mang vốn từ vựng xã hội.

Các nhà cung cấp dịch vụ thường đính kèm theo các dịch vụ mở rộng kèm theo dịch vụ chia sẻ đường link ví dụ như: cho điểm, thêm chú thích (comment) cho từng đường link một, khả năng nhập và xuất tập hợp các đường link yêu thích từ trình duyệt, từ một file dữ liệu, từ một công đồng khác.

Trên thế giới các nhà cung cấp dịch vụ này rất phổ biến ví dụ như: digg.com, del.icio.us (của Yahoo! Inc), stumbleupon.com, furl.net...



Tại sao các doanh nghiệp và cá nhân cần đến Social Bookmarking? Để thu hút lượng view lớn (traffic) các blogger và webmaster thông thường để lại đường link của blog trên một trang web có nhiều lượng truy cập và có uy tín như Facebook, Twitter…

Khi đó tùy thuộc vào độ HOT của tiêu đề, nội dung tóm tắt và hình ảnh đại diện bài viết thì link sẽ có được nhiều sự quan tâm của cộng đồng (thông qua số lượng bình chọn, bình luận) và nếu được xếp vào mục “tin hot” thì sẽ hiển thị mặc định ngay trên trang chủ và khi đó blog sẽ ngay lập tức tăng được lượng truy cập một cách đáng kinh ngạc.

## Folksonomy là gì?

Folksonomy – một biến thể phổ biến của social bookmarking, là một hệ thống dán nhãn mở được đóng góp bởi nhiều người dùng Internet, cho phép họ phân loại nội dung như các trang Web, hình ảnh trực tuyến Flickr hay video trên Youtube. Các nhãn (tag) được chọn một cách tùy ý, giúp tăng hiệu quả công cụ tìm kiếm do nội dung đã được phân loại bằng cách sử dụng những ngôn ngữ gần gũi, dễ hiểu và dễ chia sẻ.

Folksonomy là kết quả của việc gán nhãn cộng đồng (collaborative tagging – Macgregor & McCulloch, 2006) – nghĩa là các nhãn phải được chia sẻ giữa nhiều người dùng. Ví dụ một số hệ thống như Gmail cho phép người dùng gán nhãn, nhưng những nhãn đó chỉ sử dụng riêng cho một cá nhân cụ thể, thì sẽ không được xem là folksonomy.

Từ “folksonomy” được ghép lại bởi hai từ “folk” và “taxonomy”. Từ “folk” có nghĩa là bình dân, hay được tạo bởi người sử dụng bình thường thay vì các chuyên gia. Từ “taxonomy” thể hiện tính chất phân loại và sắp xếp dữ liệu. Người sáng tạo ra khái niệm “folksonomy” là Thomas Vander năm 2004.

Folksonomy thường bị nhầm lẫn với khái niệm “taxonomy” – một hệ thống phân loại phân cấp. Folksonomy gần nghĩa với khái niệm “social tagging system” – tức là một hệ thống phân loại không phân cấp dựa trên sự hội tụ của ngữ nghĩa hơn là sự chuẩn hóa.

Folksonomy được sinh ra như một đặc tính chủ chốt của Web 2.0, làm nổi bật các tính chất như tính mở, tính cộng đồng và tính tương tác. Nó được sử dụng rộng rãi trong các ứng dụng xã hội và các dịch vụ đánh dấu trang xã hội như Delicious, Flickr, và Youtube.

## Các thuộc tính quan trọng của folksonomy

Folksonomy gồm ba thuộc tính quan trọng là: tài nguyên, nhãn, và người dùng.

1. ***Phân loại tài nguyên - Resource Categorization***

Một vấn đề chủ chốt của folksonomy là sự phân loại các nguồn tài nguyên – lấy thông tin và nhóm chúng thành từng mục để tìm hiểu và phân biệt giữa các nhóm với nhau. Phân loại tài nguyên giúp phân biệt giữa khái niệm “taxonomy” và “folksonomy”.

* Taxonomy: hệ thống phân loại topo từ trên xuống dưới được nghiên cứu và sắp xếp bởi các nhà khoa học. Ví dụ trong lĩnh vực sinh học, một phân loại có cấu trúc phả hệ, nút gốc áp dụng cho tất cả các đối tượng – sinh vật (Organism); các nút bên dưới gốc là các phân loại cụ thể hơn, áp dụng cho các tập con của tập chứa các đối tượng phân loại như: giới (kingdom), ngành (phylum), lớp (class), bộ (order), họ (family)…
* Folksonomy: phát triển cấu trúc dựa trên sự đóng góp của người dùng và làm nổi bật các từ khóa phổ biến. Tính mở và linh hoạt này cho phép folksonomy dễ dàng đáp ứng với sự biến đổi hơn taxonomy, nhưng đồng thời cũng phải đối mặt với nhiều thách thức.
  1. ***Nhãn - Tags***

Nhãn (tags) cũng là một thành phần quan trọng của folksonomy. Golder và Huberman (2006) đã liệt kê các loại nhãn khác nhau dựa trên chức năng của chúng:

* Xác định nói về vấn đề gì – ví dụ, “sự kiện bóng đá U23 châu Á”
* Xác định nói về ai – ví dụ, “G.S Ngô Bảo Châu”
* Xác định nó là gì – ví dụ, “hình ảnh”
* Xác đinh chủ sở hữu – ví dụ, “nhà nước”
* Tinh chỉnh thêm thông tin các mục hỗ trợ các nhãn khác – ví dụ, số 25, 100
* Xác định tính chất hoặc đặc tính của nội dung – ví dụ, “vui vẻ”
* Các từ khóa có tính chất cá nhân cao – ví dụ, “đồ của tôi”
* Chỉ dẫn các hành động cá nhân – ví dụ, “để đọc”

Sau này, xuất hiện thêm một loại nhãn mới là nhãn địa chỉ (geotag – 2007). Nhãn địa chỉ cho phép người dùng đánh dấu vị trí của dữ liệu bằng kinh độ, vĩ độ hay tên địa điểm.

Người dùng có thói quen gán nhiều nhãn cho cùng một nội dung. Nghiên cứu chỉ ra rằng thói quen người dùng là sử dụng trước các nhãn phổ biến, sau đó sẽ thêm các nhãn cá nhân cụ thể hơn (Golder & Huberman, 2006; Hotho et al., 2006b)

Nhãn thể hiện tính chất cộng đồng của folksonomy: một nhãn khi được một người dùng tạo ra thì những người dùng khác cũng có thể biết được ý nghĩa của nó và sử dụng lại nếu muốn. Hiệu ứng xã hội sẽ giúp cải tiến và phát triển các nhãn hữu ích và loại bỏ các nhãn vô ích dựa trên tần suất nhãn được sử dụng lại và phản hồi từ toàn bộ cộng đồng mạng.

* 1. ***Người dùng - Users***

Người dùng đóng một vai trò quan trọng trong sự phát triển của folksonomy thông qua việc tìm kiếm, thu hồi, kiểm tra, đánh giá phản hồi các nhãn trong cộng đồng mà hình thành và thay đổi cấu trúc folksonomy. Marlow et al (2006) chỉ ra 6 lí do có thể ảnh hưởng tới việc gán nhãn

* Thu hồi dữ liệu trong tương lai. Ví dụ, “toread” để gán cho các nội dung sẽ cần quay lại và đọc kĩ hơn.
* Đóng góp và chia sẻ. Cung cấp dữ liệu để người khác có thể tìm kiếm.
* Thu hút sự chú ý. Sử dụng một nhãn phổ biến gán cho nội dung, tuy nhiên điều này cũng có thể bị lợi dụng để spam.
* Biểu lộ ý kiến đánh giá về một mục được đánh dấu.
* Giải trí và cạnh tranh.
* Đánh dấu sự xuất hiện của bản thân trên một resource.

# CHƯƠNG 2: THỰC TRẠNG, LỢI THẾ, VÀ THÁCH THỨC ĐỐI VỚI FOLKSONOMY

## Các dịch vụ social bookmarking nổi tiếng

Khái niệm chia sẻ trực tuyến social bookmarking xuất hiện từ những thập niên 90 của thế kỉ 20, tuy nhiên sự phát triển của social bookmarking và folksonomy bắt đầu có bước nhảy vọt từ năm 2004 cùng với sự ra đời của các dịch vụ social bookmarking.

Các dịch vụ social bookmarking là các website cho phép người dùng chia sẻ và quảng cáo các trang web, bài báo, hình ảnh hay video cá nhân. Những thành viên khác trên các website đó có thể dễ dàng xem và tìm kiếm các dịch vụ mà một người dùng cung cấp. Ngoài ra, các dịch vụ tìm kiếm lớn như Google thường hiển thị các kết quả từ các site social bookmarking này, do đó nếu một trang được đánh dấu thì đồng nghĩa với việc nó sẽ có cơ hội lớn xuất hiện trên kết quả tìm kiếm và lôi kéo một lượng “view” khổng lồ.

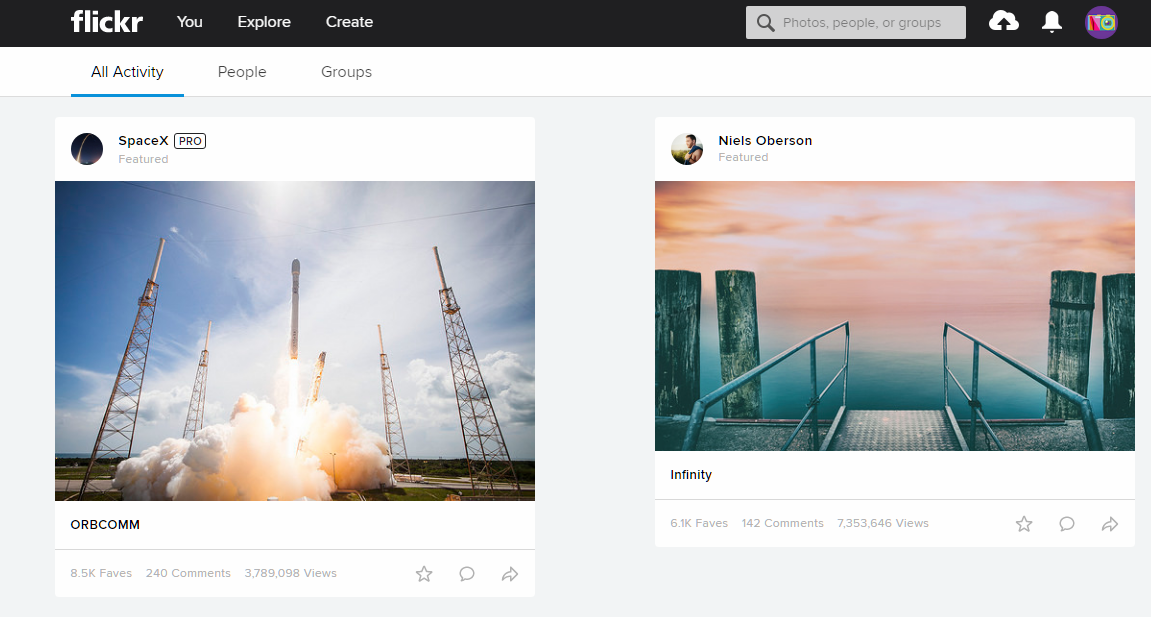
Trải qua hơn một thập kỷ cạnh tranh và hoàn thiện, một số dịch vụ có ưu thế vượt trội và chiếm lĩnh thị trường. Phần này điểm qua một số site lớn hiện nay và điểm khác biệt của chúng.

* 1. ***Flickr (https://www.flickr.com)***

Cung cấp dịch vụ lưu trữ hình ảnh và video đáng tin cậy cho bất cứ ai. Tính đến thời điểm năm 2017, flickr có tới hơn 75 triệu thành viên đăng ký từ hơn 63 quốc gia, với hơn 90 triệu lượt truy cập hàng tháng. Vào ngày cao điểm, có tới hơn 25 triêu bức ảnh được tải lên trang web này. Tổng cộng, flickr lưu trữ hơn 10 tỉ bức ảnh, và hàng tháng mỗi tháng có hơn 7 tỉ lượt sử dụng các API nó cung cấp, cùng với 1 triệu bức ảnh được chia sẻ mỗi ngày. (Tham khảo: <https://expandedramblings.com>)

Flickr cho phép người dùng gán nhãn cho nội dung của họ và cả người khác. Hình ảnh có thể để ở chế độ công khai hoặc riêng tư. Người dùng có thể tìm kiếm các bức ảnh liên quan đến cùng một nhãn bất kì, ví dụ nhãn “africa” sẽ hiển thị lên các bức ảnh hoặc video được gán nhãn “africa”.

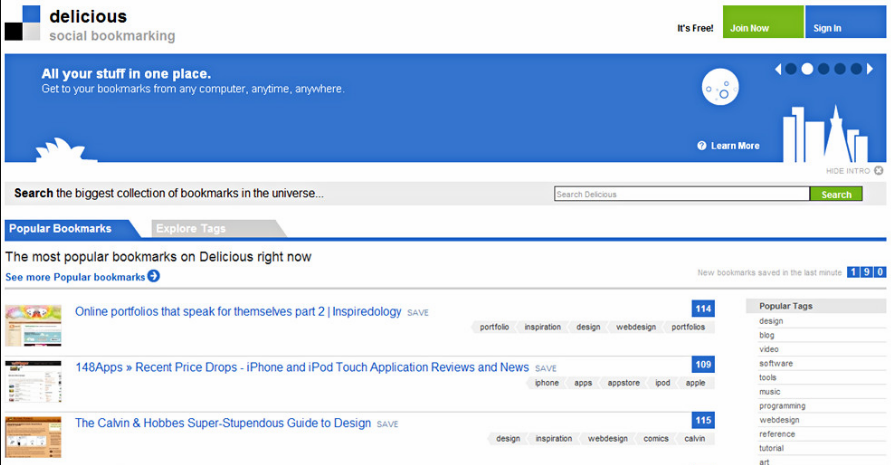
Flickr cho phép người dùng sử dụng các geotag bằng các machine tags. Sử dụng machine tags sẽ gán nhãn bức ảnh bằng kinh độ (geo:long = 123.456) và vĩ độ (geo:lat = 123.456). Người dùng sau đó có thể tìm kiếm hơn 100,000 địa điểm nổi tiếng dựa trên các nhãn này.



Tuy nhiên, flickr chỉ cho phép người dùng gán nhãn cho các bức ảnh hay video được upload lên site này, mà không thể gán nhãn cho các nội dung được lưu trữ ở địa chỉ khác. Mặc dầu Flickr cho phép người dùng gán nhãn cho nội dung của người khác, nhưng đa số bức ảnh chỉ được gán nhãn bởi chính chủ nhân của chúng (Marlow et al., 2006).

* 1. ***Delicious (delicious.com - hiện tại https://del.icio.us)***

Delicious là một trong ba nhà tiên phong trong lĩnh vực social bookmarking, cho phép lưu trữ, chia sẻ và khám phá web bookmark. Tới năm 2008, dịch vụ này có tới 5.3 triệu tài khoản và 180 triệu URL được đánh dấu. Sau nhiều lần đổi chủ sở hữu, hiện tại service của delicious đang tạm ngừng và không thể đăng ký tài khoản mới.

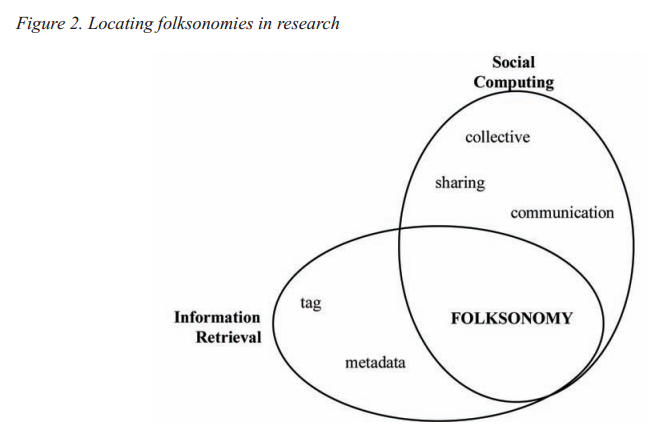


Các thành viên có thể đánh dấu và quản lý các mục trên delicious theo sở thích cá nhân và chia sẻ với thành viên khác. Hệ thống delicious là hệ thống không phân cấp, cho phép người dùng gán nhãn bằng một từ bất kỳ. Ví dụ người tìm kiếm chữ “elephant”, thì họ sẽ được liệt kê các mục được gán nhãn “elephant”, đồng thời có thể xem được tất cả các đánh dấu trang của một người khác sử dụng từ khóa “elephant”.

Delicious còn hiển thị các từ bookmark phổ biến hay các mục mới được tạo ra và các nhãn gán cho chúng. Như vậy, bằng cách truy cập dịch vụ, người dùng có thể nắm được xu hưởng hiện tại đang được ưa thích và tìm kiếm các cộng đồng có chung đam mê mới mình.

## Lợi thế của Folksonomy

Folksonomy là giao hòa giữa hai lĩnh vực trong hệ thống thông tin: tính toán xã hội (social computing) và thu hồi thông tin (information retrieval). Nó kế thừa khái niệm gán nhãn và siêu dữ liệu (metadata) từ thu hồi thông tin; tính chia sẻ và cộng đồng của tính toán xã hội.



Folksonomy có các ưu điển vượt trội so với hệ thống phân loại phân cấp đó chính là sự dễ dàng tham gia, tính phản hồi tương tác cao, và dễ dàng thu hồi thông tin của một nội dung hay đối tượng bất kì. Tuy nhiên chúng cũng dẫn đến một số thách thức sẽ được trình bày ở mục kế tiếp.

1. ***Dễ dàng tham gia sử dụng dịch vụ***

Đặc điểm đáng chú ý nhất của folksonomy là không đòi hỏi bất cứ kiến thức cao siêu nào về hệ thống để có thể sử dụng nó. Bất kì ai cũng có thể sử dụng folksonomy mà không phải hướng dẫn gì. Điều này đối lập taxonomy, để có thể sắp xếp thông tin (ví dụ như một tiến sĩ sinh vật học) thì phải học tập và nghiên cứu hàng năm trời.

1. ***Lợi ích ngay lập tức tới người dùng***

Dễ dàng tham gia là lợi ích dễ thấy nhất. Ví dụ, khi một người gán nhãn một bức ảnh trên Flickr, hoặc một nhân viên gán nhãn cho một nội dung trên mạng Intranet thì họ sẽ dễ dàng thu hồi thông tin ở lần truy cập tiếp theo. Hơn nữa, họ có thể được hưởng lợi từ việc phân loại của người khác. Dần dần, folksonomy cho phép các cá nhân tự quản lý các tài nguyên của họ, và xây dựng nên các cộng đồng thành viên có sở thích tương tự.

1. ***Tính tương tác cao***

Khi một nhãn được gán cho một đối tượng, ngay lập tức người dùng có thể thấy được các đối tượng khác đã được gán cùng nhãn đó. Sự tương tác này cho phép người dùng có thể thay đổi nhãn nếu muốn hoặc thêm các nhãn khác có liên quan.

1. ***Tính mở***

Thể hiện ở điểm người dùng có thể sử dụng bất cứ từ khóa nào mà họ mong muốn. Nhờ đó mà folksonomy thích ứng nhanh chóng với xu hướng thay đổi và sáng tạo của cộng đồng.

1. ***Duyệt và tìm kiếm***

Folksonomy rất hữu ích khi người dùng cần truy cập các thông tin thông qua các đường link gắn với các nhãn. Sử dụng folksonomy, người dùng có thể dễ dàng chuyển từ một nhãn (ví dụ “lamborghini”) sang một nhãn liên quan khác (ví dụ “Ferrari”). Nhờ vậy người dùng có thể truy cập một lượng lớn kết quả thay vì chỉ tập trung vào một từ khóa.

Đặc tính nổi bật của folksonomy là cho phép người dùng truy xuất các kết quả theo sự lựa chọn và sở thích cá nhân thông qua các từ khóa liên quan, thay vì đi theo các nội dung được định hướng bởi các chuyên gia như taxonomy hoặc ontology.

1. ***Dùng cho cá nhân, nhóm nhỏ***

Folksonomy cho phép một cá nhân hoặc một nhóm nhỏ có thể sử dụng các nhãn theo ý nghĩa của riêng họ. Ví dụ, nhãn “sometaithurts” (so meta it hurts) được gán cho một bức ảnh trên Flickr. Những thành viên khác có thể tham gia thảo luận bằng cách thêm nhãn này vào hình ảnh của họ.

## Thách thức đối với Folksonomy

Sự mơ hồ trong ngôn ngữ là thách thức lớn nhất đối với folksonomy, bao gồm: từ đa nghĩa, từ đồng âm, từ đồng nghĩa, ngôn ngữ khác nhau, dấu cách và cụm từ.

* 1. ***Từ đa nghĩa (Polysemes)***

Từ đa nghĩa là từ có nhiều nghĩa. Ví dụ trong tiếng Anh, từ “foot” có thể hiểu là “bàn chân”, hoặc là phần dưới của một thứ gì đó (“foot of the hill”), hoặc là phần thân thấp nhất của một cái cây. Tất cả chúng đều có nghĩa là nền, phần dưới, hay điểm cuối của một cái gì đó. Tuy nhiên, tùy vào ngữ cảnh cụ thể mà nó có nghĩa khác nhau. Từ đa nghĩa là nguyên nhân gây ra sự nhập nhằng trong xử lý ngôn ngữ tự nhiên, nội dung tìm kiếm càng chung chung thì càng dễ gây ra sự mơ hồ trong kết quả.

* 1. ***Từ đồng âm (Homonyms)***

Từ đồng âm là từ phát âm hoặc cấu tạo âm thanh giống nhau nhưng nghĩa hoàn toàn khác nhau. Ví dụ, trong tiếng Việt, từ “đường” có thể hiểu là “con đường” hoặc là một loại gia vị có vị ngọt. Hay câu “con ruồi đậu trên mâm xôi đậu” thì từ “đậu” mang hai nghĩa khác nhau. Sự nhầm lẫn có thể được hạn chế bằng cách thêm các từ khóa để loại bỏ bớt các kết quả không cần thiết, ví dụ “gia vị đường” hoặc “con đường”.

* 1. ***Từ đồng nghĩa (Synonyms)***

Các từ đồng nghĩa là các từ viết khác nhau nhưng có nghĩa tương tự nhau. Ví dụ, “pupil” đồng nghĩa với các từ “student”, “learner”, “scholar”, “begginer”,… Vấn đề mà từ đồng nghĩa gây ra cho người dùng là khó xác định tìm kiếm từ khóa nào trong số từ đồng nghĩa này sẽ đem lại kết quả đúng với mong muốn nhất.

* 1. ***Dạng số nhiều (plural) và đa ngôn ngữ***

Ví dụ, “rabbit” (con thỏ) hoặc “rabbits” (nhiều con thỏ) đều ám chỉ đến loài thỏ, nhưng có thể bị xem là hai từ khóa khác nhau trong folksonomy. Dẫn đến khi tìm kiếm một từ khóa thì các nội dung liên quan đến từ khóa còn lại sẽ không được hiển thị, trừ khi các dịch vụ social bookmarking có cung cấp các từ khóa liên quan.

Tương tự, đa số các hệ thống tương thích với tiếng Anh, do tính mở rộng của folksonomy nên một từ có thể được viết trong nhiều ngôn ngữ khác nhau để gán cho cùng một nội dung, ví dụ “elephant” vs “elefante” vs “con voi”. Ngoài ra, vấn đề còn xuất phát từ các cụm từ địa phương, tiếng lóng, từ mới, từ trái nghĩa.

* 1. ***Sự hỗn loạn (Entropy)***

Sự hỗn loạn của folksonomy là kết quả của việc có quá nhiều nhãn vô nghĩa phục vụ cho mục đích cá nhân duy nhất. Những từ này chỉ dùng được với một người dùng và khó hình dung với người dùng khác. Kết quả tìm kiếm với các từ khóa này thường không mang nhiều ý nghĩa hơn là nhiễu.

* 1. ***Dấu cách và cụm từ***

Hầu hết folksonomy được thiết kế một từ duy nhất không có dấu cách trống, vì dấu cách trống thường được dùng để phân cách các nhãn. Ví dụ, từ “John Lennon” thì có thể dùng các cách dán nhãn sau: loại bỏ dấu cách “johnlennon”; sử dụng chữ in hoa “JohnLennon”; sử dụng kí tự thay thế “john-lennon”. Rõ ràng điều này góp phần gây nên sự hỗn loạn trong kết quả tìm kiếm.

* 1. ***Sự biến thể phân cấp***

Sự biến thể liên quan tới các từ khóa cơ bản được dùng để phân loại nội dung. Ví dụ, một người dùng muốn gán nhãn cho một trang web về “mallard”(vịt trời), có thể lựa chọn từ khóa “bird” (chim), trong khi người khác sử dụng từ khóa “duck” (vịt) để mô tả cụ thể hơn, hoặc người khác nữa muốn sử dụng đúng từ “mallard”. Do đó, khi tìm kiếm từ khóa “bird” sẽ không trả về các tài nguyên khác chỉ được gán nhãn “duck” hoặc “mallard”. Sự khác nhau giữa trình độ và suy nghĩ của người dùng trong việc định nghĩa các nhãn ảnh hưởng tới việc liên kết các folksonomy.

* Cá nhân và tập thể

Có hai mục đích khác nhau khi gán nhãn là cho cá nhân và cho tập thể. Một số nhãn chỉ dùng cho mục đích thu hồi thông tin cho một cá nhân, trong khi một số nhãn khác được chia sẻ cho nhiều người dùng. Do đó trong quá trình thu thập nhãn trong toàn bộ hệ thống (flickr, delicious), sẽ xuất hiện nhiều nhãn được sử dụng rộng rãi trong cộng đồng, nhưng cũng có rất nhiều nhãn vô nghĩa.

* Tag spamming

Tag spamming là một thách thức không nhỏ đối với các hệ thống tìm kiếm. Nó thường được sử dụng với ý đồ không tốt đẹp là đánh lạc hướng tìm kiếm và lôi kéo sự truy cập của người dùng tới một nội dung sai bằng cách gán một nhãn phổ biến hoặc nhãn không liên quan tới một nội dung nào đó. (Koutrika et al., 2007). Koutrik xác định một vài trường hợp thực tế của tag spam. Ví dụ, một thành viên gán một nhãn đang trong top phổ biến lên các bức hình của mình dù nội dung không hề liên quan tới nhãn đó. Hoặc một doanh nghiệp có thể lợi dụng gán các trang web vào một nhãn (ví dụ, “mua xe oto”) mà loại bỏ các trang của đối thủ cạnh tranh, gây khó khăn cho việc tìm kiếm của khách hàng.

# CHƯƠNG 3: NGHIÊN CỨU GIẢI QUYẾT CÁC THÁCH THỨC

## Các giải pháp hiện nay

* 1. ***Tag Clouds***

Xây dựng một đám mây chứa các nhãn đã được sử dụng và các liên kết phổ biến. Mỗi nhãn được liên kết tới một mục đánh dấu. Một đám mây nhãn có hai đặc điểm: văn bản và vị trí từ.

* Văn bản (text): bao gồm phông chữ, kiểu, kích cỡ, màu chữ.
* Vị trí từ (word placement): bao gồm sự sắp xếp và ghép nhóm các nhãn thành một bố cục và có nghĩa nhất định. Các nhãn có thể xuất hiện theo thứ tự bất kì (ví dụ theo thứ tự từ điển, ngẫu nhiên,…).

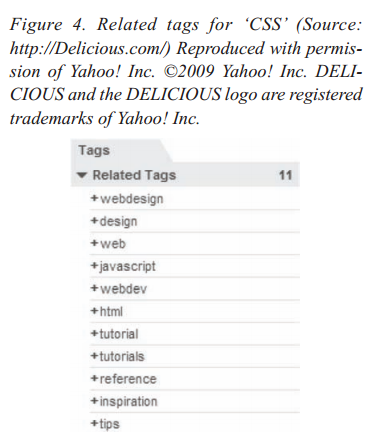
Một dạng phổ biến của tag cloud là sắp xếp nhãn theo thứ tự từ điển cùng với kích cỡ chữ tỉ lệ với mức độ được sử dụng của nhãn.



* 1. ***Related Tags***

Các dịch vụ như Delicious cung cấp một danh sách các nhãn có liên quan bằng cách sử dụng một thuật toán xác định các nhãn thường được đi kèm với một nhãn xác định trước. Điều này cho phép người dùng duyệt qua các tag liên quan và tìm kiếm các đánh dấu trang được gán cùng với từ đồng nghĩa hoặc tương đương.

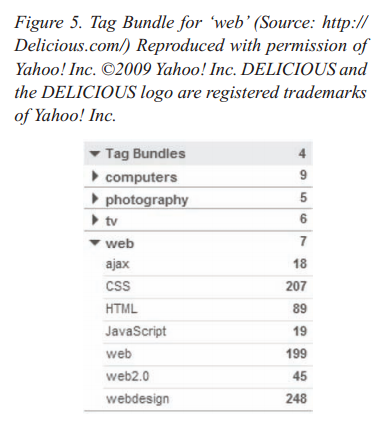
Ví dụ, từ khóa “CSS” sẽ đi kèm với danh sách các từ khóa có liên quan như “webdesign”, “design”, “web”, “javascript”,…



Cùng với việc sử dụng nhiều nhãn cho cùng một nội dung, các nhãn liên quan giúp người dùng vượt qua một số khó khăn gây ra bởi sự mơ hồ của ngôn ngữ ở chương trước.

* 1. ***Tag Bundles***

Tag bundles là nhóm các nhãn thường đi kèm với nhau thành một nhãn chung. Ví dụ, khi số lượng nhãn tăng lên, người dùng có thể lựa chọn nhóm một số nhãn thành một tag bundles. Nó giúp hình thành một cấu trúc nhãn thuận tiện cho việc quản lý và tìm kiếm. Ví dụ trên delicious, người dùng có thể gộp các nhãn “css”, “html”, “javascript” thành một nhóm gọi là “web”.



* 1. ***Tag Descriptions***

Mô tả nhãn giải quyết vấn đề từ đa nghĩa và đồng âm bằng cách hướng dẫn cụ thể cách người dùng gán nhãn với mục đích cụ thể là gì.



Mặc dù tag clouds, related tags, tag descriptions và sự tương tác tức thời khi gán nhãn có thể giúp giảm bớt các vấn đề của folksonomy, nhưng những vấn đề không thể biến mất trừ khi người dùng tuân theo quy tắc nhất định khi gán nhãn. Do đó, còn rất nhiều nghiên cứu cần được tiến hành để tìm ra các phương pháp giảm thiểu các thách thức của folksonomy.

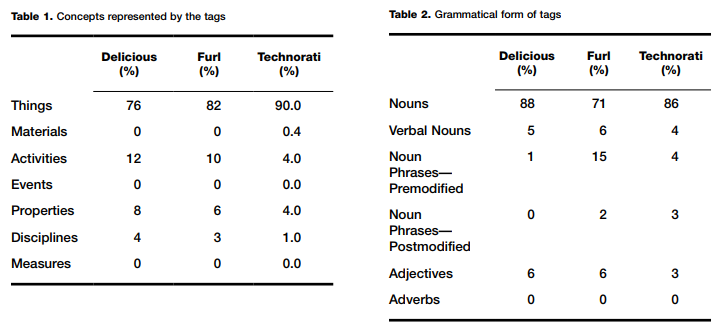
## Các hướng nghiên cứu phát triển folksonomy

Hai xu hướng chính trong việc nghiên cứu giải quyết các vấn đề của social tagging dựa trên hai thành phần tạo nên nó – thu hồi thông tin và tính toán xã hội.

* 1. ***Thu hồi thông tin***

Trong lĩnh vực thu hồi thông tin, các nghiên cứu tập trung chủ yếu vào cấu trúc và hình mẫu của folksonomy liên quan tới ngôn ngữ.

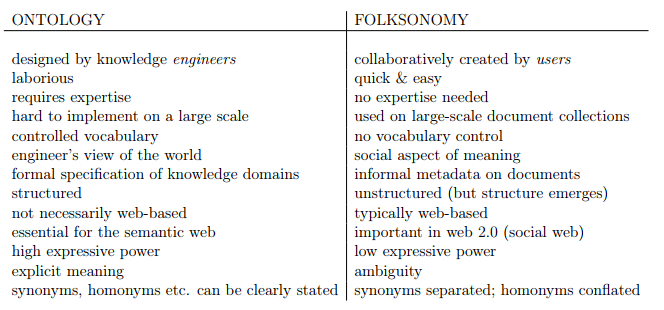
Ví dụ, Spiteri (2007) đã nghiên cứu các nhãn trên ba dịch vụ social bookmarking lớn là delicious, furl, và technorati, về các khái niệm mà tag biểu thị, dạng số ít – số nhiều danh từ, số lượng danh từ, tiếng lóng… và đưa ra các hướng dẫn cách xây dựng các nhãn sao cho có kết quả tìm kiếm tốt nhất.



Một số nghiên cứu khác tập trung vào khám phá giá trị của folksonomy trong việc tìm kiếm và tổ chức dữ liệu. Ví dụ, Sinclair & Cardew-Hall (2008) nhìn nhận sự hữu ích của tag clouds khi cần tìm kiếm các thông tin chung, trong khi đó việc tìm kiếm các thông tin cụ thể thì thường ưu tiên các trang tìm kiếm. Heymann et al. (2008) thì chứng minh social bookmarking hỗ trợ tốt cho việc tìm kiếm web (web searching) sau khi nghiên cứu đặc tính của hơn 40 triệu đánh dấu trang trên dịch vụ delicous. Ông kết luận social bookmarking có thể hiển thị một số dữ liệu mà không có ở bất kì nguồn nào khác.

* 1. ***Ontology và Folksonomy***

Sự đa dạng folksonomy cũng đồng nghĩa với sự thừa thãi, hỗn loạn gây phức tạp cho việc thu hồi thông tin. Một số nhà khoa học chỉ ra rằng, tính tương tác trong folksonomy có thể được tận dụng để hình thành nên các ontology và siêu dữ liệu. Một ontology là một hình mẫu bao gồm nhiều folksonomy dùng cho một mục đích.



Ví dụ công trình của Mika (2007) xây dựng một hình mẫu gồm 3 phần Actor – Concept – Instance để xác định một ontology bao gồm một nhóm các nhãn trên delicious.

Heymann và Garcia-Molina (2006) công bố một thuật toán cho phép chuyển đổi các nhãn thành một hệ thống điều hướng phân cấp nhãn.

* 1. ***Đối mặt với Tag spamming***

Tag Spamming ngày càng trở thành một thách thức không nhỏ đối với các nhà cung cấp dịch vụ và người dùng. Các nhà khoa học đã và đang cố gắng tìm ra những phương pháp để hạn chế thực trạng này.

Công trình của Xu et al. (2006) cho phép một tập hợp các giải pháp dựa trên đánh giá mức độ uy tín để xác định spam.

Công trình của Koutrika et al. (2007) đưa ra và đánh giá tác động của một thuật toán xếp hạng tag spam.

Các nhà cung cấp dịch vụ social bookmarking cũng áp dụng một số cơ chế anti-spamming, tuy nhiên chúng có thể gây ảnh hưởng tiêu cực tới người dùng. Ví dụ, một người dùng gán nhiều nhãn là các từ đồng nghĩa cho cùng một đối tượng có thể bị nhận dạng sai là hành vi spam.

* 1. ***Các hướng nghiên cứu khác***

Dù cho nhiều công trình đã được công bố có thể giải quyết được một số thách thức đối với folksonomy, vẫn còn rất nhiều điểm có thể cải tiến và cần giải quyết trong thực tiễn. Ví dụ, nhiều ý kiến tranh cãi hoài nghi tác dụng của folksonomy trong việc hỗ trợ tìm kiếm, thì hướng nghiên cứu các nhãn-hướng-người-dùng (user-driven tagging) mang lại lợi ích cho người dùng khả năng tổ chức nội dung theo suy nghĩ riêng, đồng thời cũng tăng chính chính xác các kết quả tìm kiếm bằng cách làm cho các nhãn dễ hình dung và có tính cộng đồng hơn là chỉ một cá nhân.

Trong lĩnh vực thu hồi thông tin, cũng có nhiều hướng nghiên cứu để tìm ra các phương pháp mới để quản lý và điều hướng các folksonomy. Ví dụ, xây dựng một hệ thống nhắc nhở các từ đồng nghĩa khi người dùng gán một nhãn, tự động kiểm tra và sửa lỗi chính tả của nhãn. Chất lượng của các folksonomy cũng có thể được tăng lên bằng cách áp dụng các hệ thống đánh giá mức điểm cho tính chính xác của nhãn.

## Kết luận

Một folksonomy được hình thành dựa trên sự đóng góp phản hồi và chọn lọc từ tất cả các người dùng trong cộng đồng thông qua các nhãn giúp chú thích, phân loại nội dung và xây dựng nên một cộng đồng con cùng chung sự quan tâm đối với nội dung đó. Folksonomy có những ưu điểm vượt trội so với những hệ thống phân loại cũ như taxonomy ở chỗ cho phép người dùng dễ dàng tham gia, trao đổi ý nghĩa các nhãn, đánh dấu để truy cập lại trong tương lai. Mặt khác, folksonomy cũng phải đối mặt với một số vấn đề dựa trên mặt ngôn ngữ (như từ đa nghĩa, đồng nghĩa, đồng âm…), sự khác nhau trong kiến thức và suy nghĩ của người dùng, và còn phải giải quyết sự hỗn loạn và nhiễu nhãn. Thông qua bài báo, tác giả đã liệt kê ra một số biện pháp đã được các nhà cung cấp dịch vụ social bookmarking triển khai để tối thiểu hóa các vấn đề (tag clouds, related tags…), đồng thời cũng nêu ra một số hướng nghiên cứu nhằm giúp cộng đồng folksonomy phát triển tốt hơn trong tương lai.