

Ontology và Web ngữ nghĩa

Lê Thanh Hương

Thông tin chung

■ Đánh giá

- Điểm quá trình: 30%
- Thi cuối kỳ: 70%

■ Website: <http://is.hut.vn/~huonglt/OnWeb>

■ Tài liệu tham khảo

- 1. D. Fensen and J. Hendler. *Spinning Semantic Web* (2003).
- 2. Grigoris Antoniou and Frank van Harmelen. *A Semantic Web Primer* (2004).
- 3. F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider. *Description Logic Handbook* (2003).

2

Web là gì

- Sáng tạo bởi Tim Berners-Lee
- Hệ thống thông tin toàn cầu dựa trên cơ sở:
 - HTTP (để truyền dữ liệu)
 - các giao thức khác: FTP, SMTP,...
 - URI (chuẩn định dạng dữ liệu)
 - URN (dữ liệu không đánh chỉ số được) và URL (dữ liệu có thể đánh chỉ số)
 - HTML (cách định dạng để biểu diễn tài liệu)
 - và các định dạng dữ liệu đa phương tiện khác: hình ảnh, âm thanh,...

Web truyền thống và các hạn chế

- là nơi máy tính thực hiện việc trình diễn và con người làm việc dịch (web ngữ pháp) và kết nối
- Nội dung và sự thể hiện hướng đến con người
- Khối lượng thông tin bùng nổ:
 - Tăng gấp đôi sau sáu tháng
- Mức độ tận dụng thông tin chưa cao
 - Các máy tìm kiếm: 25% Web thế giới
- Vấn đề Precision và Recall của tìm kiếm theo từ khóa

4

Tìm kiếm theo từ khóa

■ Từ đồng âm khác nghĩa:

□ crane:

- Sếu
- Cần cẩu

5

Tìm kiếm từ khóa

book victor hugo

Search



[Avenue Victor Hugo Bookshop -](#)

Used and new **books** with searchable catalog and newsletter.
www.avenuevictorhugobooks.com/ - 16k



[Victor Hugo: Bibliography](#)

Victor Hugo: A Companion to His Poetry, J.C. Ireson.

Hugo's Les Miserables and the **Novels** of the Grotesque, ...

www.poetry-archive.com/h/hugo_victor_bibliography.html - 21k

6

Bản chất vấn đề

■ Máy tính không hiểu “ngữ nghĩa” của thông tin.

- “Con chuột của anh bị chết rồi. Mua cho anh con khác đi.”



7

Web với Con người

The Man Who Mistook His Wife for a Hat : And Other Clinical Tales by Oliver W. Sacks

In his most extraordinary book, "one of the great clinical writers of the 20th century" (*The New York Times*) recounts the case histories of patients lost in the bizarre, apparently inescapable world of neurological disorders. Oliver Sacks's *The Man Who Mistook His Wife for a Hat* tells the stories of individuals afflicted with fantastic perceptual and intellectual aberrations: patients who have lost their memories and with them recognize people and common shout involuntary obscenities; retarded yet are gifted with un...

Our rating : ★★★★★

Find other books in : ☐ Neurology ☐ Psychology

Search books by terms :

altavista THE SEARCH COMPANY

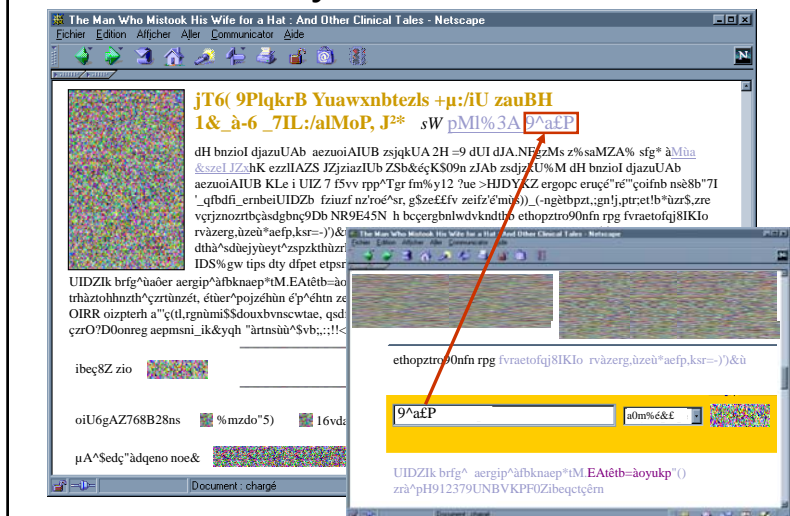
Try your search in: [Images](#) • [Video](#) • [MP3/Audio](#) • [News](#)

Search for: [Help](#) | [Custom](#)

Tools: [Shopping](#) • [E-mail](#) • [Translate](#) • [Maps](#) • [Yellow Pages](#) • [People Finder](#) • [Search](#)

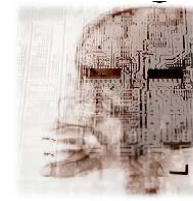
Bookmark name: [Charles](#) • [Museum](#) • [Tutor](#) • [Indonesia](#) • [Official](#) • [Menu](#) • [World](#) • [Times](#)

Web với máy tính...



Máy tính còn thiếu ???

- Tri thức
 - ☐ Mô hình chung biểu diễn tri thức
- Về cái gì
 - ☐ Trang Web
 - ☐ Thành phần của trang Web,...



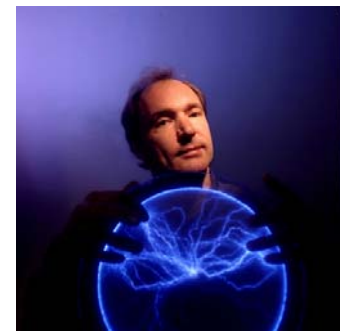
10

Giải pháp: Semantic Web

- Thay vì tạo tài liệu bằng ngôn ngữ tự nhiên
 - ☐ tạo chúng theo dạng dữ liệu máy có thể xử lý được
 - ☐ sử dụng các thuật ngữ máy có thể hiểu được để biểu diễn thông tin
- Semantic Web cần làm cho tài nguyên dễ tiếp cận hơn với các quá trình tự động bằng cách
 - ☐ Mở rộng việc đánh dấu biểu diễn với đánh dấu ngữ nghĩa
 - ☐ Sử dụng ontology để cung cấp vốn từ vựng chung cho web ngữ nghĩa
- **Ontology** mô tả các thực thể và quan hệ giữa chúng

Web ngữ nghĩa – Khái niệm & Định nghĩa

- Web ngữ nghĩa là sự mở rộng Web hiện tại, ở đó thông tin luôn được gắn với một “ngữ nghĩa” xác định – cho phép máy tính và con người cộng tác tốt hơn.



12

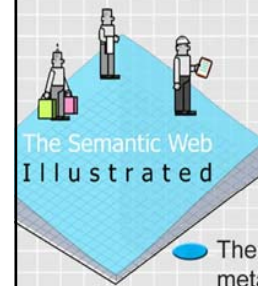
Web ngữ nghĩa – Khái niệm & Định nghĩa

- “Web ngữ nghĩa là một quan điểm với ý tưởng chủ chốt là làm cho dữ liệu trên Web được định nghĩa và liên kết theo một cách thức nào đó để chúng có thể được sử dụng bởi máy tính không chỉ với mục đích hiển thị, mà còn với mục đích tự động hóa, tích hợp và tái sử dụng dữ liệu giữa nhiều ứng dụng khác biệt.”



13

1 It's evolution not revolution



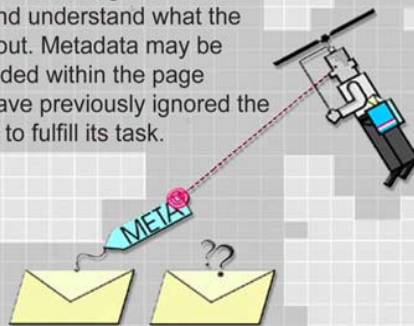
The Semantic Web leverages the content and services that already exist in the World Wide Web. By extending the World Wide Web to be more usable to robots, agents or machines, new applications are now possible.

The Semantic Web is a layer of machine understandable metadata on top of the World Wide Web

The World Wide Web is the foundation for the emerging Semantic Web

2 Add metadata. Serve to agents immediately

By attaching additional data to our existing information, agents are able to recognize and understand what the information is and what it is about. Metadata may be linked to a web page or embedded within the page itself. Where an agent would have previously ignored the information, it can now utilize it to fulfill its task.

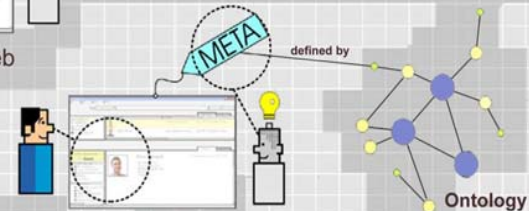


3 What was confusing, now makes perfect sense

Currently, a Web page is developed mainly for human consumption. Software agents that look at the current pages can understand very little. In the future agents will be able to understand the majority of information on the Web.



World Wide Web

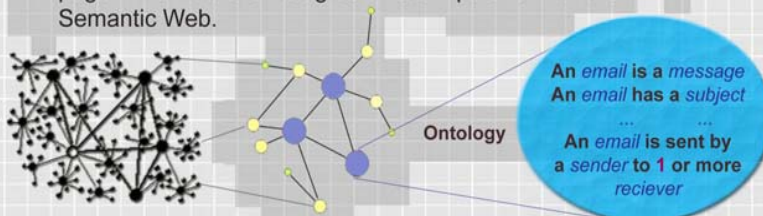


Semantic Web

Ontology

4 Ontologies give the metadata meaning.

A semantic Web page consists of RDF and linkages to global ontologies. Agents can understand the RDF associated with a page once they have crawled the web of ontologies. It is this set of ontologies or vocabularies that act as a dictionary of sorts to the agents who are attempting to understand the terms used to describe the page. This web of ontologies exists in parallel with the Semantic Web.



5 The Semantic Web isn't just about Web site either!

The Semantic Web aims to make many different types of data usable. Emails, Web sites, calendars, resumes, office documents and contacts are just some of the information resources that can be utilized by software agents if they are enhanced with metadata.



Imagine what new applications are possible once agents can tell the difference between these items!

6 How will we use the Semantic Web?

We'll use the Semantic Web using many different applications. We may have particular applications that use a specific type of information, like calendaring for example.

Alternatively, it is possible to choose an application that allows controlling a set of agents from within the same user interface. To harness the full potential of the Semantic Web, the different island of information will be used together to give new applications and greater insight.



Ontology

■ Ontology bao gồm:

- ☐ 1 bộ từ vựng mô tả các khái niệm và quan hệ giữa chúng
- ☐ Đặc tả ý nghĩa từ vựng
- ☐ Các ràng buộc mô tả các tri thức bổ sung về lĩnh vực

■ Ontology cần

- ☐ Thể hiện được hiểu biết chung về 1 lĩnh vực
- ☐ Cung cấp 1 mô hình có thể thao tác được trên máy

Tại sao cần Ontology?

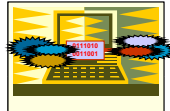
- Để các ứng dụng có thể hiểu được cùng một thông tin hay siêu thông tin
- Rộng hơn: Giải quyết bài toán trao đổi thông tin giữa các chủ thể: người – người, người – máy, máy – máy.



A short narrow tube with a small container at one end, used for smoking tobacco.



A long tube made of metal or plastic that is used to carry water or oil or gas.



A temporary section of computer memory that can link two different computer processes.

21

Định nghĩa



22

Định nghĩa « dễ hiểu »

An

- explicit
 - shared
 - formal specification
 - of the terms in the domain
 - and relations among them
1. It describes a domain
 2. A formal specification
 3. Agreed by a community
 4. No implicit information

23

Gruber, 1993] [Guarino & Giaretta, 1995] [Bachimont, 2000]

- a **logical theory** which gives an **explicit, partial account of a conceptualization** i.e. an **intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality**; the aim of ontologies is to **define which primitives, provided with their associated semantics, are necessary for knowledge representation in a given context.**

24

Ontology

- Định nghĩa bộ từ vựng chung cho các tác tử (nghĩa rộng – bao gồm cả con người) cần chia sẻ thông tin trong một lĩnh vực
- Chứa các định nghĩa (diễn dịch được bằng máy tính) về các khái niệm và quan hệ giữa chúng trong một miền ứng dụng cụ thể

25

Vai trò của ontology

- Hình thành ngôn ngữ chung để chia sẻ - tái sử dụng tri thức
 - “People can’t **share knowledge** if they do not speak a **common language**.” [Davenport & Prusak, 1998]
- Giao tiếp người – người / ứng dụng - ứng dụng tốt hơn.
- Chuẩn hóa – hình thức hóa ý nghĩa của các thuật ngữ qua các khái niệm – là nền tảng để biểu diễn tri thức.

26

Vai trò của ontology

- Chia sẻ cách hiểu chung về cấu trúc của thông tin giữa con người và tác tử phần mềm
 - Giao tiếp giữa các site thương mại điện tử
- Làm các giả định của một miền ứng dụng trở nên tường minh
 - Tránh sự cứng nhắc do định nghĩa trong mã nguồn hay các lược đồ CSDL
 - Có thể thay đổi linh hoạt

27

Phân loại ontology

- Ontology biểu diễn tri thức
 - Ontology tổng quát
- Siêu ontology
- Ontology lĩnh vực
- Ontology tác vụ
- Ontology lĩnh vực – nghiệp vụ
- Ontology ứng dụng

28

Ontology Biểu diễn tri thức

- Chứa các biểu diễn nguyên thủy được dùng để chuẩn hóa tri thức trong một mô hình
 - Frame Ontology
 - Ontology tổng quát: từ vựng liên quan đến sự vật – hiện tượng có tính khái quát cao, dùng trong nhiều lĩnh vực

29

Meta - Ontology

- Còn gọi là ontology lỗi.
- Tái sử dụng trong nhiều lĩnh vực
 - Thế nào là class, entity, relation

30

Ontology lĩnh vực

- Dùng (tái sử dụng) trong một lĩnh vực cụ thể
 - Khái niệm, quan hệ, hoạt động

31

Ontology tác vụ

- Từ vựng hệ thống hóa các thuật ngữ giải quyết một vấn đề liên quan đến tác vụ
 - Kế hoạch, mục tiêu, ràng buộc
 - ấn định, lựa chọn, phân loại

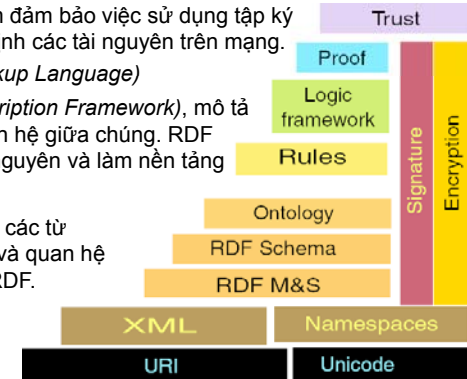
32

Ontology trong Semantic Web

- Semantic web cần sự thống nhất về các ngôn ngữ ontology trước khi chia sẻ ngữ nghĩa
- Ngôn ngữ ontology như vậy có thể dựa trên ngôn ngữ dữ liệu web
- Ngôn ngữ ontology cho semantic web có thể hình thức hóa dựa trên nền tảng của logic

Kiến trúc phân tầng của web ngữ nghĩa

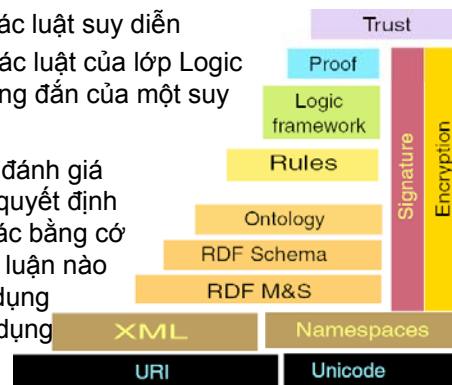
- Unicode và URI nhằm đảm bảo việc sử dụng tập ký hiệu quốc tế và xác định các tài nguyên trên mạng.
- XML (*Extensible Markup Language*)
- RDF (*Resource Description Framework*), mô tả các đối tượng và quan hệ giữa chúng. RDF gán kiểu cho các tài nguyên và làm nền tảng cho Ontology.
- RDF SCHEMA đặc tả các từ vựng mô tả tính chất và quan hệ giữa các tài nguyên RDF.



34

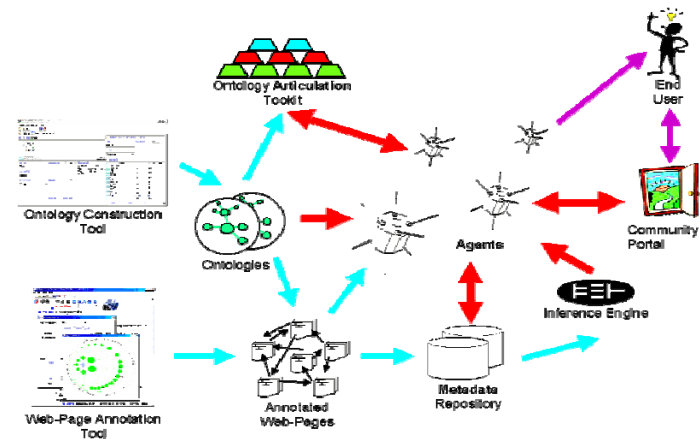
Kiến trúc phân tầng của web ngữ nghĩa

- LOGIC cung cấp các luật suy diễn
- PROOF sử dụng các luật của lớp Logic để kiểm tra tính đúng đắn của một suy diễn nào đó.
- Lớp TRUST nhằm đánh giá mức độ tin cậy và quyết định có nên tin tưởng các bằng chứng từ một kết quả suy luận nào đó hay không, sử dụng hàm lượng giá áp dụng trên một tập các thông tin



35

Khung nhìn Web ngữ nghĩa nói chung



36

Thành phần của Web ngữ nghĩa

- Ontology
- Chú thích ngữ nghĩa
- Mô tơ suy diễn
- Agent

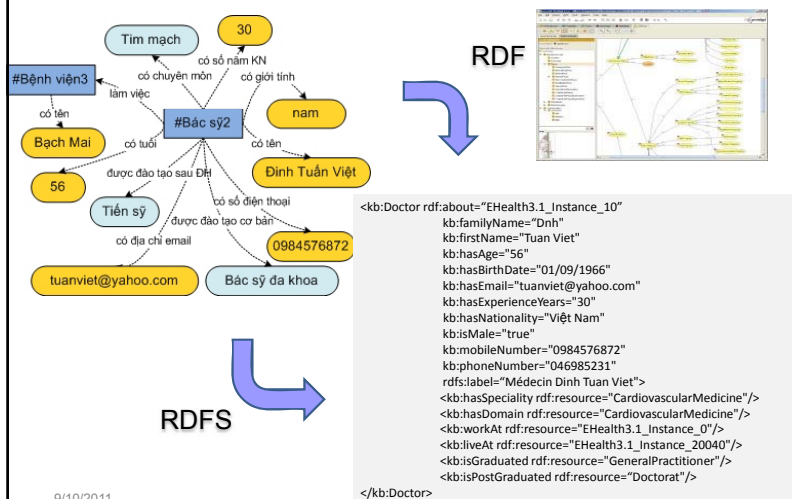
37

Chú thích ngữ nghĩa – semantic metadata

- Sử dụng ontology để diễn đạt thông tin – tri thức – dưới dạng:
 - Máy có thể hiểu được
 - Chuẩn hóa – trao đổi dễ dàng giữa các ứng dụng khác nhau
- Ví dụ: “**Bác sĩ** Nguyễn Thanh Đạt là **chuyên gia tim mạch** – **công tác tại bệnh viện Bạch Mai**. **Đã tu nghiệp 5 năm tại Đại học Y Khoa Paris**. **Chủ nhiệm dự án nghiên cứu** phương pháp phẫu thuật Van tim cho trẻ sơ sinh”

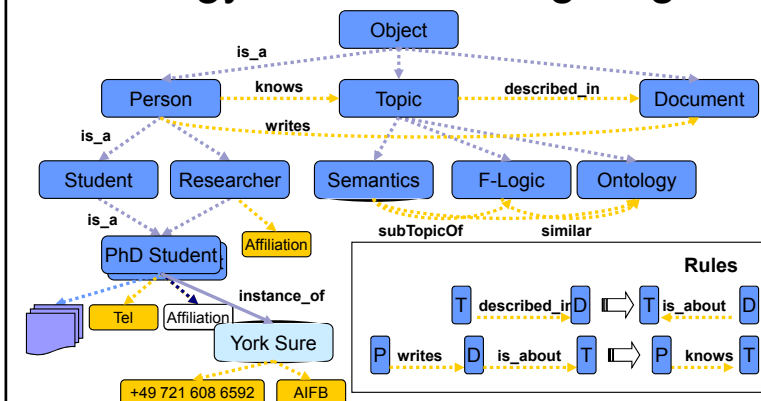
38

Một chú thích ngữ nghĩa trong lĩnh vực E-Health

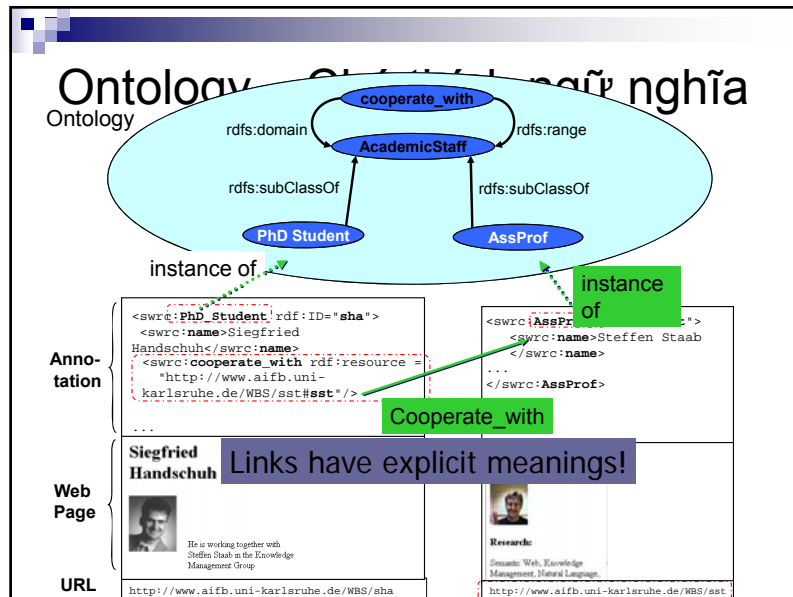


9/10/2011

Ontology – Chú thích ngữ nghĩa



40



Mô tơ tìm kiếm ngữ nghĩa

- Cho phép thực hiện hiệu quả các truy vấn ngữ nghĩa
 - ☐ Tìm kiếm trực tiếp
 - ☐ Mở rộng không gian tìm kiếm qua các luật suy diễn
- Báo cáo A1934 có được chú thích là:
 - ☐ Một instance của concept Report
 - ☐ Có quan hệ: Chủ đề liên quan là Vật liệu FZ323
 - ☐ Vật liệu Fz323 là instance của lớp vật liệu tăng ma sát trong điều kiện độ ẩm cao
- Tri thức lưu trữ:
 - ☐ Lớp xe Renault DieCast sử dụng thành phần vật liệu tăng ma sát trong điều kiện độ ẩm cao
 - ☐ Renault DieCast là instance thuộc lớp xe đua F1, etc

42

Mô tơ tìm kiếm ngữ nghĩa

- Hướng tiếp cận xây dựng
 - ☐ Đồ thị quan niệm (Conceptual Graph)
 - Corese, JENA
 - ☐ Logic Mô tả (Description Logic):
 - Racer, Pellet, KAON, FacT++

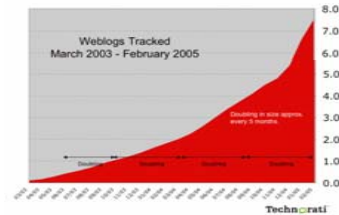
43

Xã hội hoá web

- Các phần mềm xã hội hóa
 - ☐ Cho phép con người lập lịch hẹn, hợp tác qua kết nối máy tính hoặc cộng đồng online.
 - ☐ Các công cụ cho liên kết trực tuyến
 - Gửi thông điệp trực tuyến
 - Chat
 - Diễn đàn
 - Blogs, Weblogs
 - Wikis
 - Mạng xã hội
 - V.v....

Blog/weblog

- Sự bùng nổ của blog
 - Tăng gấp đôi/5 months
- Các dạng blog
 - photoblog, travelblog, ...



Mạng xã hội

- Mạng xã hội là cấu trúc xã hội tạo bởi các nút – là các cá nhân hoặc các tổ chức
- Lý thuyết về mạng xã hội coi quan hệ xã hội là các nút và liên kết. Nút là các cá thể trong mạng, liên kết là quan hệ giữa các cá thể.
- Các mạng xã hội
 - Freindster, Yahoo 360, ...



Web trong tương lai

