


 STI · INNSBRUCK


## Finding Media on the Web: A Child's Play?

Katharina Siorpaes  
katharina.siorpaes@sti2.at

© Copyright 2012 STI · INNSBRUCK www.sti.innsbruck.at

**STI Innsbruck: Semantic  
Technology Institute** 

- Research institute at the University of Innsbruck
- Prof. Dr. Dieter Fensel
- Semantic technologies, Ontologies, Semantic Web services, semantics for Service Oriented Architectures etc.
- Roughly 65 members from 20 different countries
- Involved in 20 running European and national research projects
- 21 closed ones




www.sti.innsbruck.at

**Projekte: eine Auswahl** 


- myOntology: Web 2.0 Ontologie Editor, Österreichisches Projekt (FFG, FIT-IT Semantic Systems), 2006 – 2009. 
- ACTIVE: Knowledge powered enterprise, EU Projekt, IP, FP 7, 12 Partner, 2008 – 2011. 
- INSEMTIVES: Incentives for Semantic Content Creation, EU Projekt, STREP, FP 7, 7 Partner, 2009 – 2012. 

www.sti.innsbruck.at

**Agenda** 


1. Einführung
2. Web 2.0
3. Web 3.0, Semantic Web & Ontologien
4. OntoGame: Spiele mit einem Nutzen

www.sti.innsbruck.at

**Einführung** 

- **Knowledge Management** wird in Zeiten der Informationsflut immer wichtiger.
- Nach wie vor können Suchmaschinen nur nach Keywords, aber **nicht nach Zusammenhängen** suchen.
- Sie können nicht **verstehen**, was auf Webseiten steht oder worum es in Videos geht.

www.sti.innsbruck.at

**Einführung** 

- Computer können Aufgaben automatisch ausführen, wenn sie diese Aufgaben vorher gelernt haben. Für dieses Training wird der Mensch benötigt.
- Das Problem ist, dass es aufwendig ist, Computer zu „trainieren“.
- Dieses Problem ist als „Knowledge Acquisition Bottleneck“ bekannt.
- **Menschliche Intelligenz ist für viele Knowledge Management Tasks notwendig!**

www.sti.innsbruck.at

**Einführung**

- „Kinderleichte“ Ansätze adaptieren, um das Wissen zu akquirieren.
- Vielen Nutzern ermöglichen, beizutragen.
- Anwendungen im Web 2.0 waren höchst erfolgreich und haben immense Datenmengen generiert.
- Viele dieser Paradigmen werden jetzt auf andere Gebiete angewandt.

**Web X.0**

**Web 2.0** (Mitmach Web)

Web 2.0 ist ein Begriff für eine Reihe interaktiver und kollaborativer Systeme des Internets.

Was war Web2.0? Web2.0 ist das Web!

From: Frank Schönfeld: Social Software, 2nd Life und Wikinomics, Innsbruck 2008

**Was ist Web 2.0? „Definition“ von O'Reilly.**

Web 1.0	Web 2.0	Neuerung
DoubleClick	Google AdSense	<b>personalisiert</b>
Ofoto	Flickr	tagging, community
Britannica Online	Wikipedia	community, free content
Webseiten publishing	<b>blogging</b>	dialog
CMS	participation	
directories	<b>wikis</b>	flexibility, freedom
taxonomy	<b>tagging</b>	community
	folksonomy	


Consumers → Prosumers

**Was ist Web 2.0? - Beispiele**

- Gmail
- Google Notebooks (Kollaboratives Notepad im Web)
- Wikis
- Wikipedia: Größtes Lexikon der Welt, Top 30 Webseite, 100 Sprachen
- Del.icio.us (Social Tagging für Bookmarks)
- Flickr (Photo Sharing und Tagging)
- Blogs, RSS, Blogger.com
- Programmableweb.com: 150 web-APIs


## Blogs

- Einfachere Benutzungsschnittstelle zum Aktualisieren von Inhalten
- Einfache Organisation der Inhalte
- Einfachere Nutzung der Inhalte
- Öffentliche Kommentare
- Technisch wenig neues
- Sozial: Kollaborativ (einzeln aber stark vernetzt)**



## Wikis

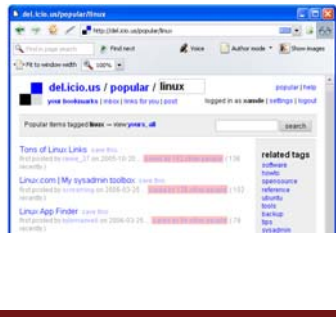
- Wiki → „schnell“ erfunden von Ward Cunningham
- Eine Sammlung von HTML Seiten: lesen und editieren
- Bekanntestes und größtes Wiki: Wikipedia (MediaWiki)
  - Aber: Auch zahlreicher Einsatz in Intranets (z.B. in unserer Gruppe)
- Probleme werden sozial statt technisch gelöst
- Flexible Struktur
- Hintergrundalgorithmen + menschliche Intelligenz
- Technisch: Wenig neues
- Sozial: Kollaborativ (kein Besitztum an Inhalten)**



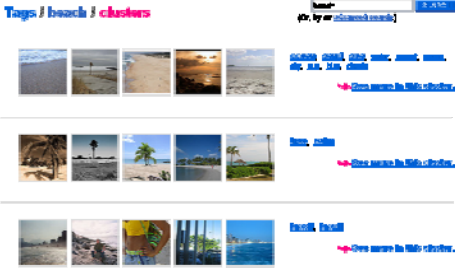
## Soziales Tagging (Vorschlagwortung)

- Idee: Auszeichnen von Inhalten mit frei vergebenen Stichwörtern
- Weg von Ordern hin zu Organisation durch Tags
- Neu: Gute (simple) Benutzungsschnittstellen für Tagging und Tag-basierte Suche
- Erste Schritte zum Semantic Web?
- Technisch: Benutzerschnittstellen
- Sozial: Kollaborativ (eigene Inhalte, gemeinsame Tags)**

## Kollaboratives Tagging

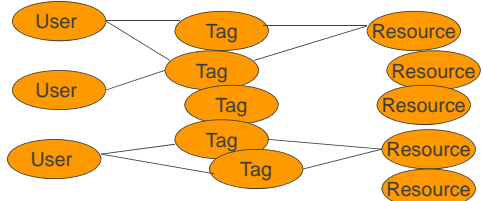


## Tagging: Flickr.com



## Folksonomies

Die aus Tagging resultierenden Daten, Wissensstrukturen



Mary tags [www.wikipedia.org](http://www.wikipedia.org) with wiki wikipedia encyclopedia  
 Bob tags [www.wikipedia.org](http://www.wikipedia.org) with wiki web2.0 encyclopedia knowledge

## Was ist Web 2.0? Trends bei Webanwendungen



- Technische Entwicklung
  - Web-Benutzungsschnittstellen werden flüssiger (AJAX)
  - Desktop verlagert sich ins Web (GMail, Google Notebooks, AJAX)
- Soziale Entwicklung
  - Kollektiv von Nutzern schafft einen Mehrwert (Wiki, Tagging)
  - Freie Inhalte setzen sich durch (Lizenzen)
  - Aufmerksamkeit wird monetarisiert (Text-Ads)
  - Mehrwert-Webseiten durch Rekombination (Mash-Ups, RSS)

www.sti-innsbruck.at

19

## Powerful Web 2.0: eine Auswahl



www.sti-innsbruck.at

20

## Powerful Web 2.0: eine Auswahl



- File Sharing:
  - Flickr (Bilder)
  - YouTube (Videos)
  - Wikipedia (Online Enzyklopädie)
  - Blogs
  - Open Source Community (Linux)
- Soziale Webseiten und Kommunikation:
  - Facebook
  - LastFM
  - Skype
  - StudiVZ
  - LinkedIn, Xing
- Dateimanagement
  - Tagging
- Offene Systeme: APIs, teilweisen open source erlauben Weiterentwicklung durch Benutzer

www.sti-innsbruck.at

21

## facebook



### Internetplattform zur Bildung von sozialen Netzwerken

- 2004 gegründet
- **64 Millionen** aktive Mitglieder
- Durchschnittl. **250,000 Neuregistrierungen** am Tag seit Jan. 2007
- Mehr als die Hälfte der Benutzer sind nicht mehr im College
- Mehr als **65 Milliarden page views** im Monat
- Mehr als die Hälfte der Mitglieder nutzen Facebook **täglich**
- Durchschnittl. Dauer: 20 Minuten
- Geschätzter Marktwert **15 Milliarden Dollar**

www.sti-innsbruck.at

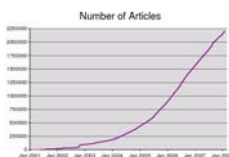
22

## Wikipedia



Freie Online Enzyklopädie

- **2,214,717 Artikel** (englische Wikipedia)
- **6,383,758 registrierte Benutzer**
- Klavere Mechanismen kombiniert mit menschlicher Intelligenz
- Hohe Qualität der Artikel
- Unkontrolliert und offen



www.sti-innsbruck.at

23

## YouTube

Broadcast Yourself™



Videportal: kostenloser Up- und Download von Videos

- In einem einzigen Monat wuchs die Anzahl der Videos um 20% auf **6.1 Millionen**
- 45 Terabyte Videos
- **1.73 Milliarden Video Views**
- Im ersten Jahr verbrachten alle User insgesamt 9,305 Jahre damit, YouTube Videos anzusehen
- Google kaufte YouTube letztes Jahr für ca. **1.6 Milliarden Dollar**

www.sti-innsbruck.at

24

## Mechanical Turk

The screenshot shows the Amazon Mechanical Turk interface. The 'Make Money' section highlights that users can work on HITs (Human Intelligence Tasks) from home, control their own work hours, and get paid for doing good work. The 'Get Results' section explains that users can complete HITs, earn money, and get paid for doing good work. It also mentions that users can work on a global, on-demand, 24/7 workforce and that their work is verified by Amazon.

## Intrinsische Motivationen

- **Altruismus**
- **Eigennutzen** (Tags: Organisation, Wiederverwendbarkeit)
- **Anerkennung** in einer Gemeinschaft
- **Zugehörigkeit** zu einer Gemeinschaft, ein gemeinsames Ziel
- **Autonomie, Willensfreiheit**
- **Aufmerksamkeit**
- **Selbstdarstellung** (Facebook)
- **Soziale Komponente** (Menschen kennenlernen)

(Kuznetsov, 2004; Marlow et al., 2006)

Unternehmen machen sich dieses Potential bereits zu Nutze ...

## Semantic Web: Das große Versprechen.

The social graph just connects people  
The semantic graph connects *everything*

And it uses richer semantics to enable:

- Better search
- More targeted ads
- Smarter collaboration
- Deeper integration
- Richer content
- Better personalization

From: Frank Schönfeld: Social Software, 2nd life und Wikinomics, Innsbruck 2008  
Nova Spivack: [www.mindtheplanet.net](http://www.mindtheplanet.net) quoted under CC.

## Semantic Web

- Idee: Webseiten angereichert mit maschinenlesbaren Annotationen basierend auf Ontologien
  - Suche über **eindeutige Konzepte** statt mehrdeutige Stichworte
    - Bsp: *Bank (Finanzinstitut)* statt „Bank“
  - **Strukturierte Suche** statt nur Stichwortsuche
    - Bsp: <„liegt in, Europa“> statt „Stadt: Europa“
  - **Inferenz** findet auch nicht explizites Wissen
    - Bsp: <Karlsruhe, Deutschland> liegt in, Deutschland> und <Karlsruhe, liegt in, Europa> liegt in, Europa>
- Stand der Technik:
  - Austauschformate RDF, OWL sind W3C-Standards (HTML, CSS, XML)
  - RDF & OWL Tools incl. Inferenz vorhanden
- Beispiel:
  - Friend-of-a-Friend (FOAF) „semantische Visitenkarte“

Based on Völkl, Vrandečić and colleagues.

## Semantic Web + Web 2.0 = Web 3.0?

	Web 2.0	Web 3.0
Tagging	<ul style="list-style-type: none"> <li>- annotieren Tags</li> <li>- Singular/Plural-Problem</li> <li>- Synonyme</li> <li>- Keinerlei Intelligenz</li> </ul>	<ul style="list-style-type: none"> <li>- annotieren mit Konzepten</li> <li>- Inferenz (Tag „Hund“ folget Tag „Tier“)</li> </ul>
Rekombination von Daten verschiedener Quellen	<ul style="list-style-type: none"> <li>• Mash-Ups vorab von Hand programmiert</li> </ul>	<ul style="list-style-type: none"> <li>• Spontan durch End-Nutzer</li> </ul>
Suche	<ul style="list-style-type: none"> <li>• Stichwortsuche oder Tag-Suche <i>findet</i> Dokumente</li> </ul>	<ul style="list-style-type: none"> <li>• Strukturierte Suche kombiniert Daten und <i>erzeugt</i> Dokumente</li> </ul>
Zeithorizont	<ul style="list-style-type: none"> <li>• 2004 - 2007</li> </ul>	<ul style="list-style-type: none"> <li>• 2007 - 2010</li> </ul>

Based on Völkl, Vrandečić and colleagues.

## Web 3.0 Ansätze

- **Automatisches Extrahieren** von Wissen auf der großen (und frei verfügbaren) Menge von Daten, die durch Web 2.0 generiert werden
- **Integration und Wiederverwendung** von Wissen (Yahoo Pipes)
- **Motivation** von Benutzern für die Generierung von semantischen Inhalten durch Anwendung von Web 2.0 Paradigmen
- **Erstellen von Semantik** als Nebeneffekt von Arbeitsprozessen (Semantic Wikis)

## Wissensmodellierung

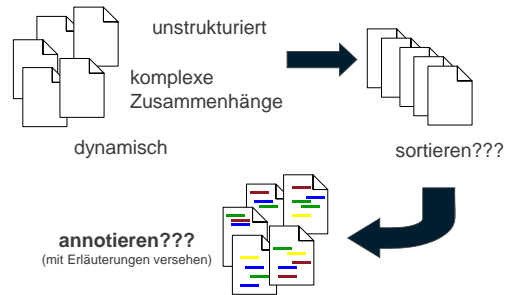


- Warum wollen wir Wissen modellieren?
- Beispiele aus dem Alltag:
  - Suche von bestehendem Wissen in grossen Mengen von Dokumenten
  - Match making: Finden von passenden Objekten anhand von verschiedensten Eigenschaften

www.htw-innsbruck.at

21

## Dokumentensuche



www.htw-innsbruck.at

22

## Komplexes Matchmaking



www.htw-innsbruck.at

23

## Knowledge Modelling



- Worum geht es bei der Wissensmodellierung:
  - Darstellung der Problemdomäne
  - Strukturierung der Darstellung, Schematisierung
- Verständnis der Probleme, der Domäne, der Zusammenhänge und Begriffe
  - Wissensmodellierung ist der erste Schritt zu Interoperabilität
  - Knowledge Sharing

www.htw-innsbruck.at

24

## Einführung und Motivation



- In Zahlen ausgedrückt:
  - Informationsverarbeiter verbringen mehr als **27 Stunden die Woche** mit dem Suchen, Sammeln und Analysieren von Information
  - Dabei werden **3.5 Stunden/Woche** verschwendet mit dem Suchen von Information, die gar nicht existiert, und
  - weitere **3 Stunden** mit dem redundanten Erstellen von schon existierenden Informationen...

www.htw-innsbruck.at

25

## Ontologie: Definition



Philosophie: die Lehre des Seins, der Essenz der Dinge  
Informatik:

"An ontology defines the **basic terms and relations** comprising the vocabulary of a **topic area** as well as the **rules for combining terms and relations** to define extensions to the vocabulary" (Neches, 1991)

"An ontology is an **explicit specification** of a **conceptualization** of a **domain**" (Gruber, 1993)

"Ontologies are defined as a **formal specification** of a **shared conceptualization** of a domain" (Borst, 1997)

www.htw-innsbruck.at

26

## Typische Eigenschaften von Ontologien

- Modelliert eine Domäne anhand von Konzepten und Eigenschaften, Relationen und Bedingungen
- Definiert
  - Einen gemeinsamen Wortschatz
  - Die Bedeutung von Begriffen
  - Wie Begriffe in Beziehung zueinander stehen
- Besteht aus
  - einer Konzeptualisierung und einer Implementierung
  - Modellierungsprimitiven
    - Klassen/Konzepte/Entitäten, Objekte/Instanzen, Attribute, Beziehungen/Relationen

www.innsbruck.at

## Anwendungen

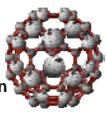
- Semantic MediaWiki
- myOntology
- OntoGame

www.innsbruck.at

## Semantic Wikipedia: Vorteile

### Strukturiertes Wissen wird exportiert (im RDF-Standard)


- Neue Web 2.0 Anwendungen möglich
- Wiederverwendung von Wissen über Sprachgrenzen hinweg
- Aggregierte Suche über mehrere Seiten
- Qualität: Finden von Fehlern und Widersprüchen
  - Hat jedes Land eine Hauptstadt?
  - Wird jede Person geboren bevor sie stirbt?
  - Passen die Geburtstagsjahreslisten zu den Geburtstagen der Personen?
  - Passt die Bevölkerungsdichte zu Bevölkerung und Fläche?



Based on Völk, Vrandečić and colleagues.

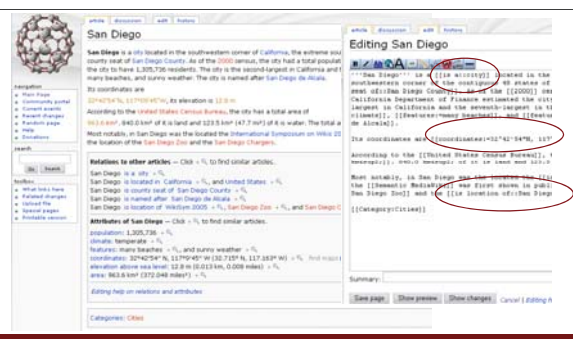
## Semantic Wiki = Wiki + Semantic Web

- Semantic MediaWiki: Erweiterung der MediaWiki-Software (Technik der Wikipedia)
- Syntaxerweiterung erlaubt getypte Links
  - Bisher: ... liegt im Süden von [[Deutschland]] ...
  - Neu: ... liegt im Süden von [[liegt in:Deutschland]] ...
- Syntaxerweiterung erlaubt annotieren von Werten
  - Bisher: ... hat eine Bevölkerung von 280,000 Einwohnern. ...
  - Neu: ... hat eine Bevölkerung von [[Bevölkerung:=280000]] Einwohnern.



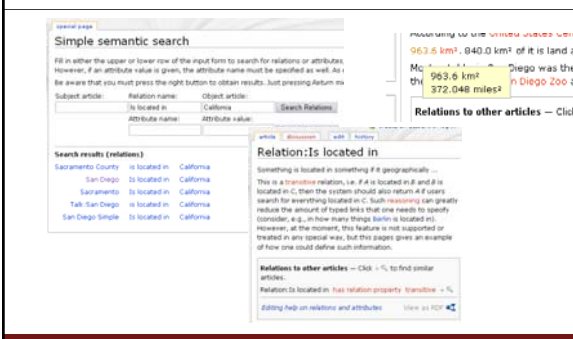
www.innsbruck.at

## Semantic MediaWiki




Based on Völk, Vrandečić and colleagues.

## Was liegt in Kalifornien?



Based on Völk, Vrandečić and colleagues.

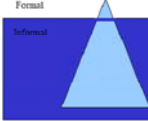
**myOntology** 

- Kollaboratives Erstellen von Ontologien
- Ein Tool, um Fachleuten und Ontologieexperten eine effiziente Kollaboration zu ermöglichen
- myOntology bedient sich des Web 2.0 Paradigmas

**Sprache ist zentraler Bestandteil des Entwicklungsprozesses**

- **Lexikale Ressourcen** sind in den Bau von Ontologien integriert
- Der **Diskurs**, der zu einer Definition führte ist verfügbar sein


All das ermöglicht besseres Verständnis und verhindert Missverständnis



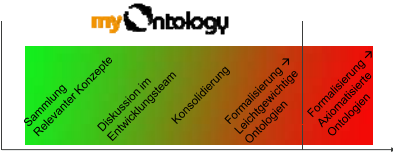
**Reichhaltige Medien**

*Verständlichkeit*

- Vewenden einer Ontologie setzt Verständnis voraus
  - Konstrukte (subClassOf)
  - Beabsichtigte Bedeutung, Semantik
- Die Serialisierung der Ontologie ist nicht der einzige Kommunikationskanal
- myOntology verwendet reichhaltige Medien (Text, Bilder, Videos, etc.)




**Verschiedene Reifegrade von Ontologien**



Uschold and King, 1999  
Braun et al., 2007

**myOntology Philosophie**



1. **Kollaboration** von Fachleuten und Ontologieexperten
2. In den ersten Phasen (bis zu **leichtgewichtigen Ontologien**)
3. Hohe **Benutzerfreundlichkeit**
4. Integration und Wiederverwendung von **Wissen im Web** (Web 2.0: Folksonomies, Flickr, YouTube, Wikipedia, etc.)
5. **Hintergrundintelligenz** unterstützt Entwicklungsteam

**myOntology Technologien**

- **Automatische Extraktion** von leichtgewichtigen Wissensstrukturen aus **Folksonomies**
- Verwendung von YouTube, Flickr, Wikipedia und Synonymen zur **reichhaltigen Beschreibung** von Konzepten, Attributen, Relationen und Ontologien
- **Export** der Ontologien in gängigen Standards
- Entdecken von **Inkonsistenzen im Hintergrund**