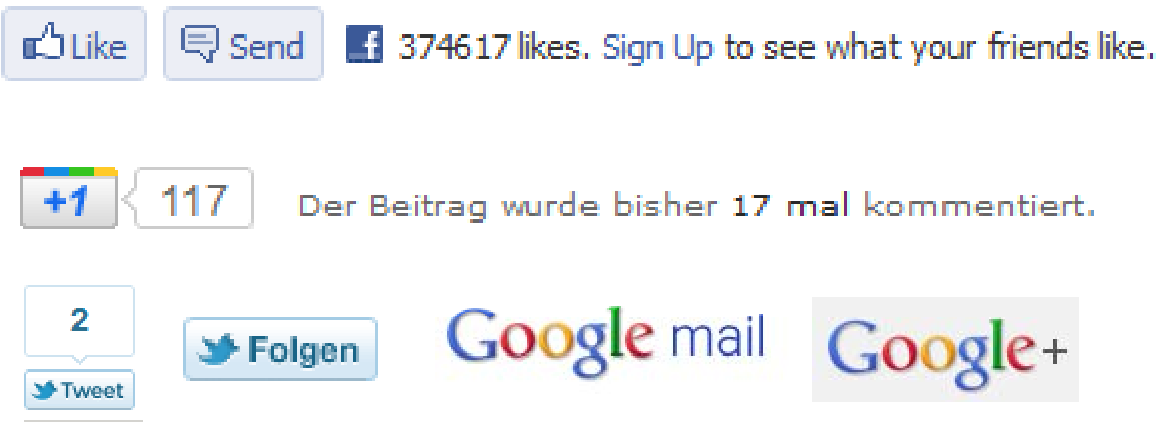
Recommender Systems and the next-generation Web

New opportunities

- **More types of information available**
- **More willingness of users to contribute**
- **New application areas**
  - Friends, pictures, movies, tags, bookmarks



Web 2.0

- **Web users connect via social networks**
  - Publish their demographic characteristics and preferences
  - Actively provide and annotate resources such as images or videos
  - Share their knowledge in community platforms
- **New types of public information spaces**
  - Web logs (blogs)
  - Wikis
  - Platforms for sharing multimedia resources
- **New capabilities of Web 2.0 greatly influence the field of recommender systems**

RS and the Social Web

- **The Web 2.0 / Social Web**
  - Facebook, Twitter, Flickr, …
  - People actively contribute information and participate in social networks
- **Impact on recommender systems**
  - More information about user's and items available
    - Demographic information about users
    - Friendship relationships
    - Tags on resources
  - New application fields for RS technology
    - Recommend friends, resources (pictures, videos), or even tags to users

  => Requires the development of new algorithms

  => Currently, many papers published on this topic

Trust-aware recommender systems (TARS)

- **Trust in recommender systems**
  - Get users to believe that the recommendations made by the system are correct and fair
  - Assess the "trustworthiness" of users to discover and avoid attacks on recommender systems
  - Trust relationships between users (our focus)
- **Trust-enhanced nearest-neighbor recommender systems**
  - Exploit trust networks to improve the system performance
  - The accuracy of the recommendations can be increased
  - Alleviate the cold-start problem
  - Improve on the user coverage

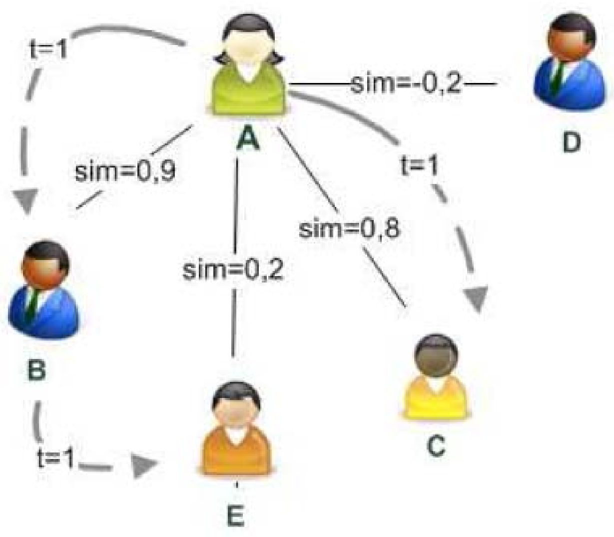
Trust-aware recommender systems (TARS)

- **Explicit trust statements between users**
  - Can be expressed on some social web platforms (epinions.com)
  - Could be derived from relationships on social platforms
  - Trust is a multi-faceted, complex concept
  - Goes however beyond an "implicit" trust notion based on rating similarity

- **Exploiting trust information in RS**
  - To improve accuracy (neighborhood selection)
  - To increase coverage
  - Could be used to make RS robust against attacks

TARS (Massa & Avesani 2007)

- **Input**
  - Rating matrix
  - Explicit trust network (ratings between 0 – no trust, and 1 – full trust)
- **Prediction**
  - Based on usual weighted combination of ratings of the nearest neighbors
  - Similarity of neighbors is however based on the trust value



Note

- Assume standard Pearson CF with min. 3 peers and similarity-threshold = 0.5
- No recommendation for A possible
- However, assuming that trust is transitive and 3 trusted users are sufficient, then the rating of E could be used
- Good for cold-start situations
- Limit transitivity

Trust metrics and effectiveness

- **Experiments on an Epinions.com dataset**
- **Effectiveness of simple algorithms**
  – Simple algorithms such as "always predict value 5" or "always predict the mean rating value of a user" (Many 5-star ratings in the dataset)
  – Predict average rating of items, good results for cold-start users. However, for controversial items CF outperforms simple algorithms
- **Using direct trust only**
  – Uses only the opinions of users for which an explicit trust statement is available
  – Works well for cold-start users, niche items and opinionated users (have a high standard deviation in their ratings),
  – Best method with respect to mean absolute user error (MAUE)
    MAUE: compute the mean error for each user and then average these user errors over all the users. Errors of cold-start users are as influential as errors for heavy rater
  – However coverage is below CF

Trust metrics and effectiveness (cont.)

- **Trust propagation**
  - Increasing propagation distance leads to an increase in rating coverage but decreases prediction accuracy
- **Hybrids**
  - Such a combination quite intuitively leads to increased coverage but the performance did not increase

RS, social networks and trust

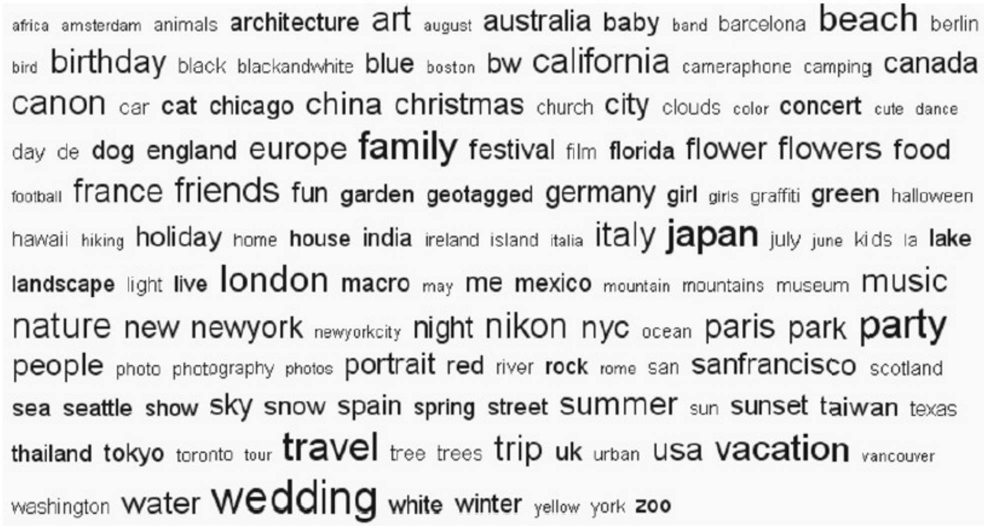
- **Hybrids**
  - Information from various sources might be combined to generate personalized information services (Hess et al. 2006), i.e. combine trust networks of researchers and visibility of scientific papers
- **Implicit trust**
  - One will ask friends who have similar tastes for a recommendation.
  - Trustworthiness is measured by how often a user has been a reliable predictor in the past (Massa and Avesani 2007)
- **Recommending new friends**
  - Another form of cold-start problem
  - Many of today's social web platforms aim to increase the connectivity of their members by suggesting other users as friends, e.g. "close a trust triangle" by similarity measures

Folksonomies

- **Folk taxonomies**
  - Users add tags to resources (such as images)
  - Tags can describe different aspects of a resource such as content, genre but also personal impressions such as *boring*
  - Folksonomies are based on freely-used keywords (e.g. on flickr.com)
  - Not as formal as ontologies, but more easy to acquire
- **Semantic Web approaches**
  - Formal, defined, and machine-processible annotations
  - Formal ontologies have the advantages of preciseness and definedness, they are hard to acquire
- **Recommender systems and folksonomies**
  - Exploit the information of how items are tagged by the community
  - Recommend tags to users

Folksonomies and content-based methods

- **Recommendations based on tag clouds**



- **Linguistic methods for tag-based recommendation**
  - merge tags assigned by users to descriptions in special slots (Gemmis et al. 2008)

Recommendations based on tag clouds

- $n_k(u,r)$ .. number of movies annotated by keyword k, assigned a rating r by user u
- $T_{u,r}$ .. tuples $\langle k, n_k \rangle$ where k is a keyword and n_k is the number of how often k was assigned by u to movies with rating r
- Given a user u, a movie m^* and a rating r^* the appropriateness of r^* is:

$$\sigma(u,m^*,r^*) = \sum_{\{\langle k,n_k \rangle \in T_{u,r^*} \mid k \in K_{m^*}\}} \frac{n_k(u,r^*)}{\log(N_k)}$$

N_k .. global frequency of keyword k

K_{m^*} .. set of keywords associated to m^*

$\log(N_k)$ .. the usual weighting factor for term frequencies

Recommendations based on tag clouds (cont.)

Weighted average for all possible rating values R

$$\overline{\sigma}(u, m^*) = \frac{1}{S(u, m^*)} \sum_{r \in R} r \times \sigma(u, m^*, r)$$

where the normalization factor is:

$$S(u, m^*) = \sum_{r \in R} \sigma(u, m^*, r)$$

Recommendations based on tag clouds (cont.)

$\overline{r}(m^*)$ .. average rating of users who have rated m^*

The weighted estimated rating value of a movie m^* of user u is

$$\sigma^*(u, m^*) = 0.5\,\overline{r}(m^*) + 0.5\,\overline{\sigma}(u, m^*)$$

Does well for average ratings, improvements possible for extreme ratings

Linguistic methods for tag-based recommendation

(Gemmis et al. 2008)

- **Items are described by static slots, e.g. title, painter**
- **In addition so called dynamic slots**
  ***SocialTags(I)* and *PersonalTags(U,I)* are added**
  I is an item, *U* is a user
  - *SocialTags(I):* tags added to *I*
  - *PersonalTags(U,I):* tags added by user *U* to *I*
  - Words in slots are replaced by synsets (synonymy set) exploiting WORDNET
  - Word sense disambiguation methods are applied
  - Slots contain a set of synsets (semantic tags)
- **Finally, a Bayesian approach is applied for predicting the user rating exploiting the values of the slots**

Folksonomies and collaborative filtering methods

- **Tag-enhanced "classical " collaborative filtering methods**
  - View tags as additional information for discovering similarities between users and items
  - For example, Tso-Sutter et al. (2008) viewed tags as additional attributes providing background knowledge
- **Tag-based collaborative filtering and item retrieval**
  - Social ranking (Zanardi and Capra 2008), a method that aims to determine a list of potentially interesting items in the context of a user query
  - Social ranking aims to overcome this problem by applying traditional CF ideas in a new way
  - Use user and tag similarities to retrieve a ranked list of items for a given user query

Tag-enhanced collaborative filtering

- **Difference to content-boosted CF**
  - Tags/keywords are not "global" annotations, but local for a user
- **Possible approach, a combined, tag-aware CF method**
  - Remember, in user-based CF
    - Similarity of users is used to make recommendations
    - Here, view tags as additional items (0/1 rating, if user used a tag or not); thus similarity is also influenced by tags
  - Likewise, in item-based CF, view tags as additional users (1, if item was labeled with a tag)
- **Predictions**
  - Combine user-based and item-based predictions in a weighted approach
  - Experiments show that only combination of both helps to improve accuracy

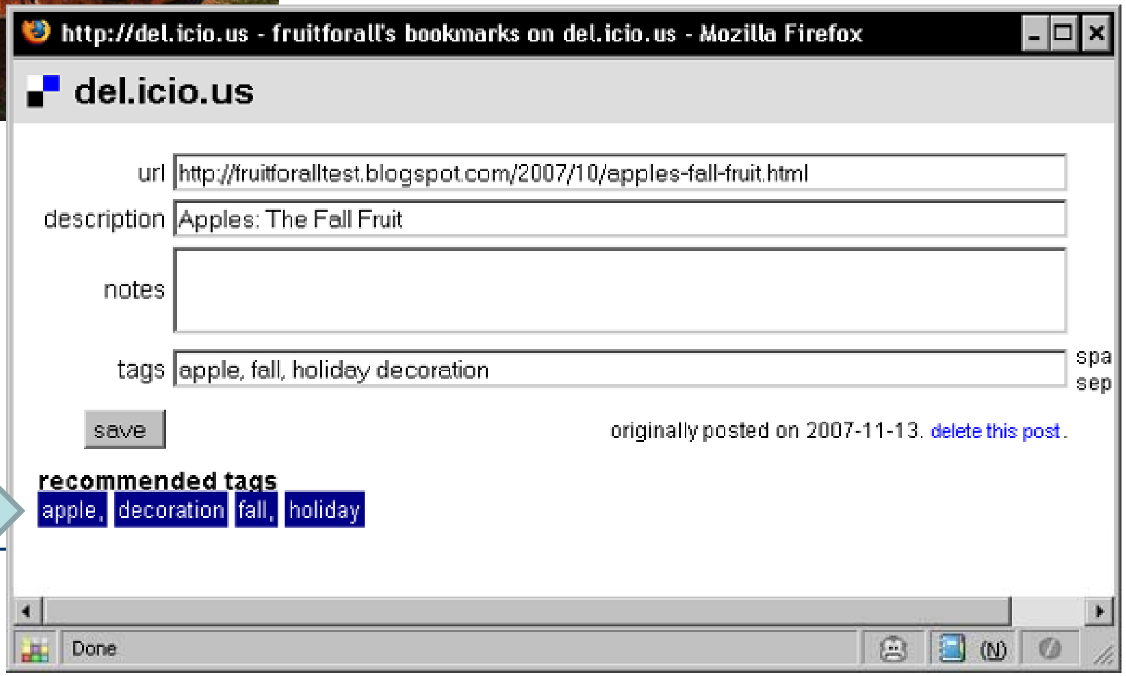
Tag-based CF and item retrieval

- **Item retrieval in Web 2.0 applications**
  - Often based on overlap of query terms and item tags
  - Insufficient for retrieving the "long tail" of items
    - Users may use different terms in their annotations
      - Think of possible tags of a car, "Volkswagen", "beetle", "red", "cool"…
- **One approach, Social Ranking**
  - Use CF methods to retrieve ranked list of items for given query
    - Compute user and tag similarities (e.g., based on co-occurrence)
  - Two-phase retrieval
    - Extend user query with similar tags (improves coverage)
    - Rank items based on
      - Relevance of tags to the query
      - Similarity of taggers to the current user
  - Leads to measurably better coverage and long-tail retrieval

Recommending tags

Tags

outback • australia • silverton • broken hill • vw • beetle • volkswagen • car • sky • rusty



Recommending tags

- **Remember, users annotate items very differently**
- **RS technology can be employed to help users finding appropriate tags**
  - Possible approach
    - Derive two-dimensional projections of the $\langle User, Tag, Resource\rangle$ relation eliminating either tags or resources
    - Determine ***k*** nearest neighbors of a user ***u*** based on one projection
    - Tag ***t*** for an item ***I*** and user ***u*** is rated by counting the usage of tag ***t*** for item ***I*** by the nearest neighbors of ***u*** weighted by the similarity of the neighbors to ***u***
    - Recommend the top ***n*** tags
  - Evaluation
    - Similarity based on User-Tag projection is better than User-Resource projection
    - Always better than "most-popular (by resource)"-strategy
- **FolkRank**
  - view folksonomy as graph and apply PageRank idea

Recommending content in participatory media

- **Second-generation web, participatory media**
  - Users contribute the content
  - Exploit information if the active user trusts the content providing person.
- **(Seth et al. 2008)**
  - Credibility of messages depend on credibility of authors which depends on topics and the active user and the opinion of her friends
  - Messages are labeled with their authors
  - Users assign a supposed credibility to messages
  - Users are explicitly connected with their "friends"
  - Every user can declare a list of topics in which he or she is interested, i.e. topic specific networks can be generated
  - Bayesian model predicts if the active user will find a new message credible

Recommending content in participatory media (cont.)

- **(Guy et al. 2009)**
  - Differentiates users in **familiar** and **similar** users w.r.t. active user
  - Familiar score depends on organizational charts, direct connections in social networks, tagging of persons, co-authorship of content
  - Similarity score depends on co-usage of tags, co-bookmarking the same web page, co-commenting the same blog entry
  - Recommendations based on similarity scores and familiarity scores were compared
  - Explanations in terms of persons who are similar/familiar were given
  - Recommendations based on familiarity scores outperformed similarity scores (user classified the recommended items as interesting, not interesting, already known)
  - Effect could be caused by persuasion
  - Explanations caused an increase of classifying items as interesting

Ontological filtering

- **Semantic Web community**
  - Describe web resources by languages that can be interpreted by software systems
  - Match the information need of users by exploiting machine interpretable information, e.g. OWL
  - Formulate a domain ontology
- **Apply ontology to improve recommender systems**
  - Knowledge-based techniques such as simple inheritance taxonomies and logical description
  - These recommender systems are actually hybrid systems
  - The aim is to leverage their capabilities by knowledge-based methods

Augmentation of filtering

- **Augmentation of filtering by taxonomies**
  - Hierarchical ontology
  - "sport" is a parent of "soccer" and a grandparent of "world soccer tournaments"
  - Use item profile and user profile to annotate news items and let users directly express interests
- **Augmentation of filtering by attributes**
  - Attributes used to characterize items
  - In the movie domain, attributes are genre, actors, director, and name
  - Use semantic information about items (e.g. genere, actor, etc.) to compute similarities between items
  - Combine semantic similarity and rated similarity to predict user ratings

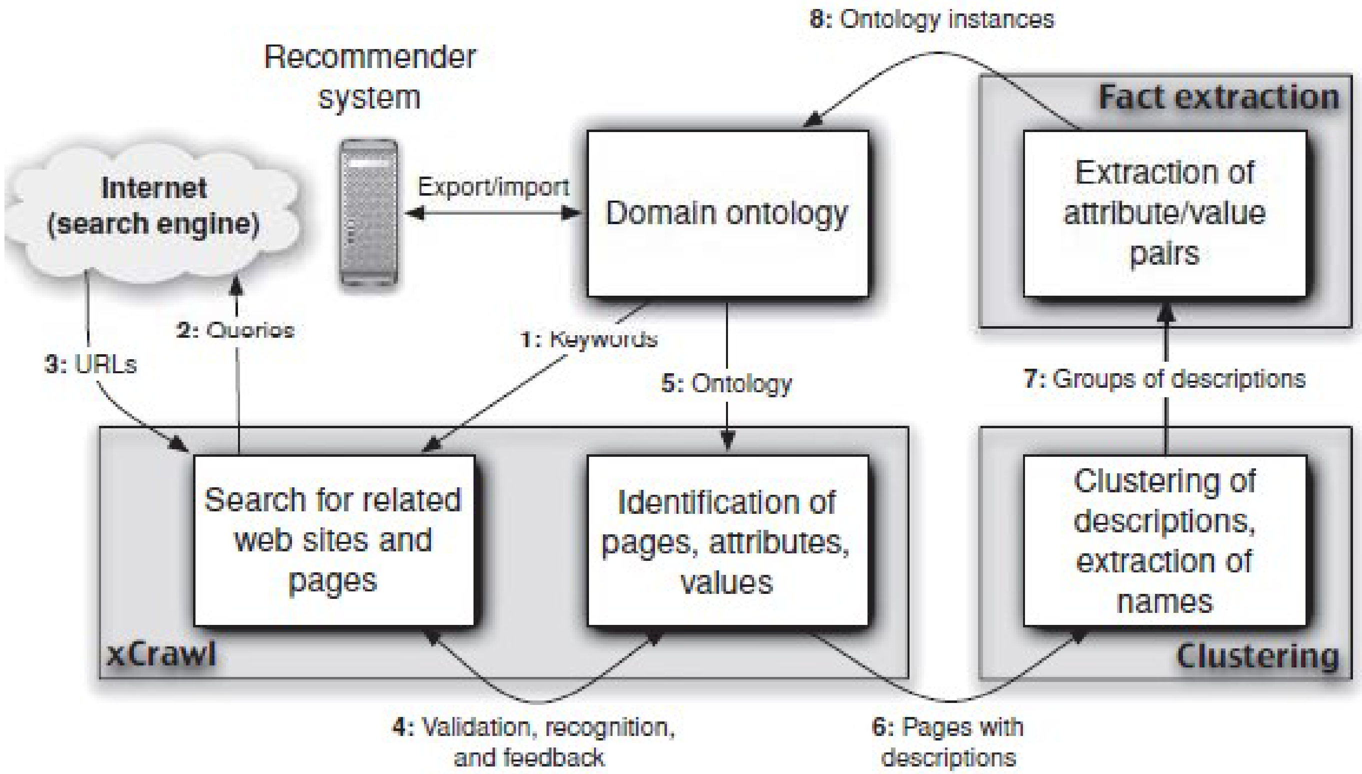
Example for filtering by taxonomies (Maidel et al. 2008)

- **Given**
  - Item profile: set of concepts associated to items
  - User profile: set of concepts associated to users
  - Taxonomy of concepts (sub-super concept hierarchy, e.g. soccer is a sport)
- **Compute matching scores between user and item concepts**
  - Various cases: perfect match, parent/child and grandparent/grandchild match e.g. user is interested in sports, item is a member of soccer items
  - Each match has a score depending on matching case
  - Compute item/user match depending on the weights of the concepts of the active user and the matching score of user concepts and item concepts
- **Evaluation**
  - Without concept taxonomy the quality of recommendations drops significantly
  - If user explicitly states the interest in concepts, quality improves significantly

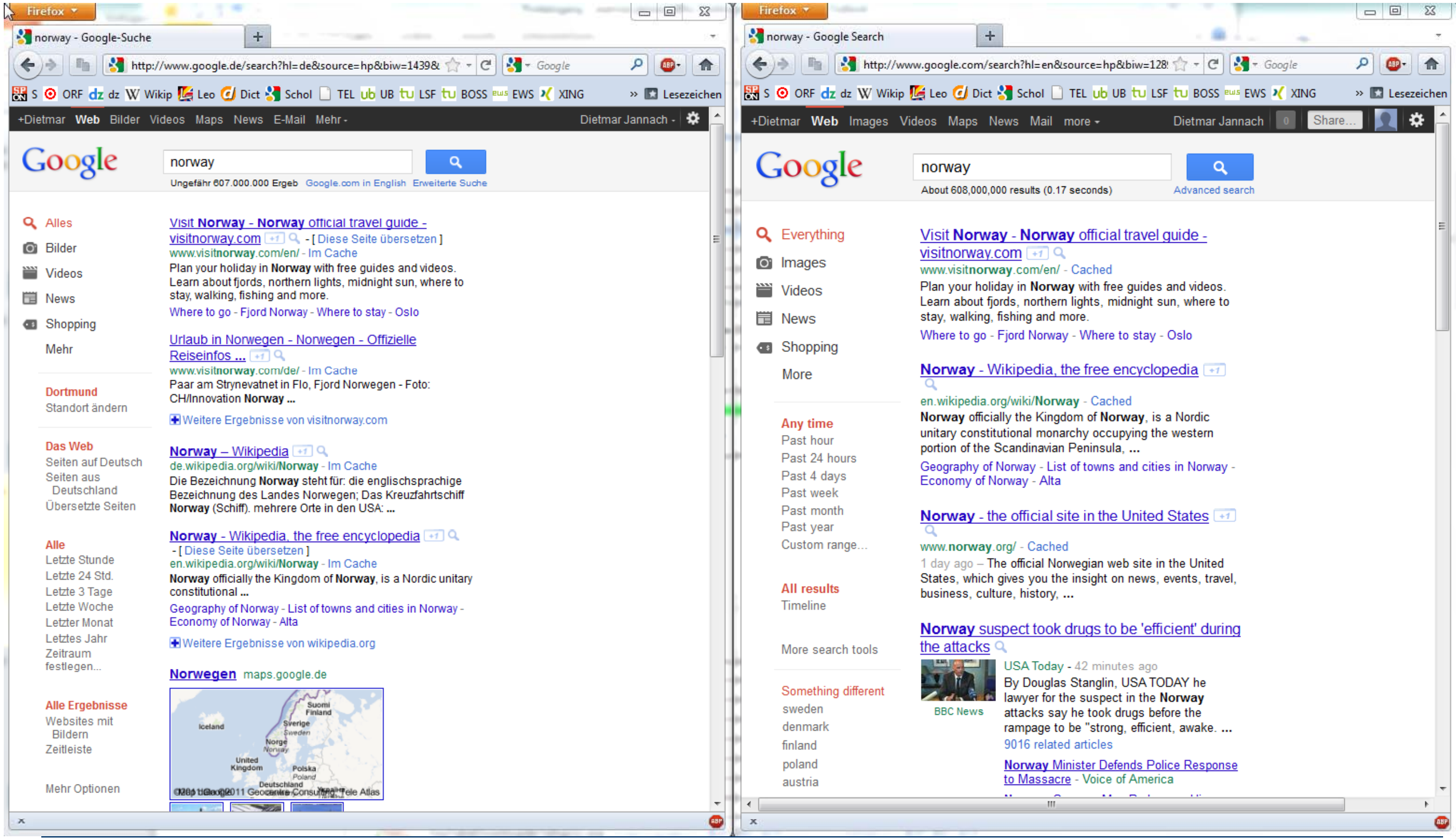
Extracting semantics from the web

- **Semantic information can provide valuable means for improving recommendations**
  - Where does this information come from?
  - How costly and reliable is the acquisition process?
- **Approaches to generate semantic information**
  - Humans are providing semantics by annotating content and by declaring logical sentences
  - Develop software systems that are able to generate semantics with little or no human intervention (particularly attractive)

AllRight system (Jannach et al. 2009)



Discussion – The Filter Bubble



Summary

- Opportunities, current methods, and realizations of Web 2.0
- Semantic Web for recommender systems
- Exploit additional information to contribute more trustworthy and qualitative enhanced recommendations
- Both Web 2.0 and the Semantic Web in combination not only drive new technologies but have huge impacts on society regarding the communication and interaction patterns of humans
- Recommendations shape the users' behavior in Web++

Literature

- **[Gemmis et al. 2008]** Integrating tags in a semantic content-based recommender, Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08), ACM, Lausanne, Switzerland, 2008, pp. 163–170.
- **[Guy et al. 2009]** Personalized recommendation of social software items based on social relations, Proceedings of the 2009 ACM Conference on Recommender Systems (RecSys '09) (New Your, USA), ACM Press, 2009.
- **[Hess et al. 2006]** Trust-enhanced visibility for personalized document recommendations, Proceedings of the 2006 ACMSymposium on Applied Computing (SAC '06) (Dijon, France) (Hisham Haddad, ed.), ACM, 2006, pp. 1865–1869.
- **[Maidel et al. 2008]** Evaluation of an ontology-content based filtering method for a personalized newspaper, Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08) (Lausanne, Switzerland), ACM Press, 2008.
- **[Massa and Avesani 2007] Trust-aware recommender systems**, Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys '07) (Minneapolis, MN, USA), ACM Press, 2007.
- **[Seth et al. 2008]** A subjective credibility model for participatory media, Workshop Intelligent Techniques for Web Personalization and Recommender Systems (ITWP) at AAAI '08 (Chicago), AAAI Press, 2008, pp. 66–77.
- **[Tso-Sutter et al. 2008]** Tag-aware recommender systems by fusion of collaborative filtering algorithms, Proceedings of the 2008 ACM Symposium on Applied Computing (SAC'08) (Fortaleza, Ceara, Brazil), ACM, 2008, pp. 1995–1999.
- **[Zanardi and Capra 2008]** Social ranking: Uncovering relevant content using tag-based recommender systems, Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08) (Lausanne, Switzerland), ACM Press, 2008, pp. 51–58.