

CONTENT RECOMMENDATION METHOD USING SOCIAL NETWORK ANALYSIS

Jong-Soo Sohn, Un-Bong Bae, In-Jeong Chung

Department of Computer and Information Science
Korea University
Republic of Korea

Abstract

- Content Recommendation Systems (CRS)
 - Important aspect of social network services and computing
- Traditional CRMs
 - Cannot be used in the Web 3.0 environment
- In this paper, we propose a novel way
 - Recommend high-quality web content
 - Using degree of centrality and Term Frequency-Inverse Document Frequency (TF-IDF)
 - TF-IDF and degree of centrality
 - Collected RDF Site Summary (RSS) and Friend-Of-A-Friend (FOAF)
 - Results from the implementation
 - More appropriate and reliable contents than traditional CRSs.
 - Reflects the importance of the role of content creators.

Introduction

- Web 2.0, the role of content creators
 - Sifted from service providers to service users
 - Variety of social network services have become very popular
 - Participate and interact between users
- Social network services
- Users can create and release content
 - Based on their experiences and knowledge using Web 2.0
 - Experiences
 - Quality of content
 - Content recommended through social networks

Introduction

- The Internet and its accompanying information technology
 - Developing at an astonishing rate
 - Volume of content and media
 - Rapid development have shown
 - “Deluge of information”
 - Difficult obtaining items that they really want and need
- CRS
 - Search and recommend suitable items
 - Traditional content recommendation methods
 - Content-based recommendation method
 - Collaborative recommendation method
 - Hybrid recommendation method

Introduction

- Proposed method in this paper
 - FOAF and RSS
 - Designed to overcome the limitations of the existing recommendation systems.
 - FOAF
 - To analyze social network services
 - RSS
 - To analyze contents
 - XML-based vocabularies offered in the Web 2.0 environment
- Four steps
 - Query entry and data collection
 - FOAF and RSS offered in the Web 2.0 environment are collected and stored in a database.
 - RSS content are analyzed using the TF-IDF technique
 - To measure the degree of similarity between query and various contents
 - Analyze the social network in terms of degree of centrality
 - Rank individual content
 - Based on the outcome of the evaluations
 - In the third and fourth steps
 - Recommend content

Background Technology

- FOAF
 - RDF/OWL vocabulary based on XML ontology
 - Users' information and activities
 - Easily share users' information with the online community
- RSS
 - RDF/XML-based standard web content publication language
 - Method of content representation
 - Unlike the traditional method
 - Automatic delivery of newly registered content
 - From a mobile application, other websites, or a user's computer

Background Technology

- TF-IDF

- Indicator that represents the degree of relevance of document
- Consists
 - Term Frequency (TF)
 - Inverse Document Frequency (IDF)
- Computed as the product of the value of TF and IDF
- Rank the results retrieved by a search engine

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad idf_i = \log \frac{|D|}{|\{d : t_i \in d\}|}$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i$$

Related Works

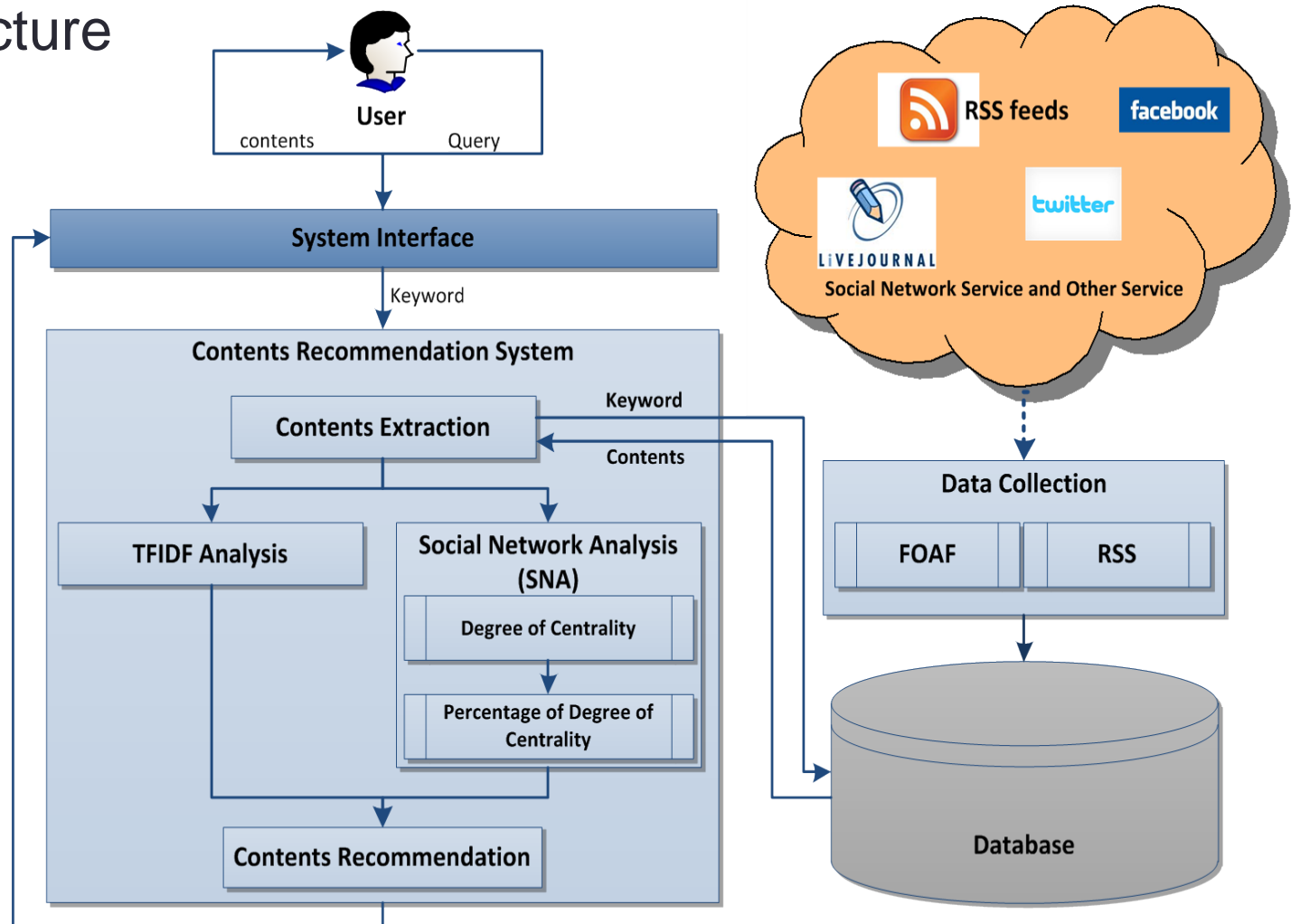
- Content recommendation systems
 - “Deluge of information”
 - Getting more difficult to identify the content that users want
- Studies of CRSs
 - Recommending necessary information resources to users
 - Conducted from diverse angles
 - CRSs is being widely used to recommend multimedia content
 - Such as music, movies and videos, and books and news articles

Related works

- Content-based recommendation method
 - Based on information retrieval
 - Contents of items and user preferences
 - Directly analyzed in order to recommend the resulting items
- Collaborative recommendation method
 - Classify users according to the similarity
 - Preferences and evaluation of other users
 - Websites that provide content such as movies or music
 - Users evaluate their chosen content and make recommendations
- Hybrid recommendation method
 - Combine techniques or other methods
 - Data mining, DNA algorithm, and artificial neural networks
 - Content-based recommendation, collaborative recommendation

Suggested CRS

- Architecture



Suggested CRS

- Recommendation steps
 - 1) Data gathering and query entry
 - 2) RSS Contents Analysis Using TF-IDF
 - 3) Social Network Analysis Using FOAF
 - 4) Content Recommendation

Suggested CRS

- Step1: Data gathering and query entry
 - Collect FOAF and RSS with query entry
 - Query refers to the keyword that users wish to recommend

```
- <item>
  <guid isPermaLink="true">http://adafrog.livejournal.com/337514.html</guid>
  <pubDate>Tue, 14 Sep 2010 17:22:49 GMT</pubDate>
  <title>AO3 Invite</title>
  <link>http://adafrog.livejournal.com/337514.html</link>
  <description>Actually still have it. Leave a comment with your email, and I'll give
  screened.</description>
  <comments>http://adafrog.livejournal.com/337514.html</comments>
  <category>ao3</category>
  <category>flist</category>
  <lj:security>public</lj:security>
  <lj:reply-count>0</lj:reply-count>
</item>
```

RSS

```
- <foaf:Person>
  <foaf:nick>clannuisnigh</foaf:nick>
  <foaf:name>clannuisnigh</foaf:name>
  <foaf:openid rdf:resource="http://clannuisnigh.livejournal.com/" />
  <ya:country dc:title="UK" rdf:resource="http://www.livejournal.com/directory.bml?o
  <ya:city dc:title="London" rdf:resource="http://www.livejournal.com/directory.bml?
  opt_sort=ut&s_loc=1&loc_cn=UK&loc_st=&loc_ci=London" />
  <foaf:dateOfBirth>1983-02-19</foaf:dateOfBirth>
+ <foaf:page>
  <foaf:icqChatID />
  <foaf:aimChatID />
  <foaf:jabberID />
  <foaf:msnChatID />
  <foaf:yahooChatID />
+ <ya:blogActivity>
  <foaf:weblog rdf:resource="http://clannuisnigh.livejournal.com/" lj:dateCreated="200
  03-26T12:17:12" />
- <foaf:knows>
  - <foaf:Person>
    <foaf:nick>elfflame</foaf:nick>
    <foaf:member_name>elfflame</foaf:member_name>
    <foaf:tagLine>Little Fall of Rain</foaf:tagLine>
    <foaf:image>http://l-userpic.livejournal.com/33239411/3938525</foaf:image>
    <rdfs:seeAlso rdf:resource="http://elfflame.livejournal.com/data/foaf" />
    <foaf:weblog rdf:resource="http://elfflame.livejournal.com/" />
  </foaf:Person>
  </foaf:knows>
+ <foaf:knows>
+ <foaf:knows>
```

FOAF

Suggested CRS

- Step2: RSS Contents analysis using TF-IDF
 - Search is conducted for content
 - Close relationship with the query
 - Assuming that
 - A user entered a certain query that a user wishes to search for
 - Measuring the degree of similarity
 - Between the entered query and the content

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad idf_i = \log \frac{|D|}{|\{d : t_i \in d\}|}$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i$$

Suggested CRS

- Step 3: Social network analysis using FOAF
 - Degree of centrality
 - Number of direct connections a node has
 - Index that measures the activity of a node
 - To analyze the degree of importance of the particular user

$$C_D(v) = \text{Degree}(v) / n - 1$$

- Rate of the degree of centrality among users related to the query

$$P_{i,m} = \frac{C_D(v_{i,m})}{\sum_{r=1}^n C_D(v_{i,r})} \times 100$$

Suggested CRS

- Step4: Content recommendation
 - determine the rankings of content
 - For user input queries using the values of $P_{i,m}$ and TF-IDF
 - Multiply the degree of content importance
 - TF-IDF
 - Degree of centrality

$$SR_{i,j,m} = tfidf_{i,j} \times P_{i,m}$$

- $SR_{i,j,m}$ reflects
 - Degree of importance of contents for the keyword l
 - Corresponding users in the social network
- We can rank the contents using $SR_{i,j,m}$

Case Study

- Environments
 - User ID: H0
 - Query: computer
 - Amount of collected content: 1,000
 - Amount of collected FOAF: 500

Contents	User	Number of words	Frequency of query appearance
C1	H1	100	5
C2	H2	80	3
C3	H3	200	8
C4	H4	50	3
C5	H5	60	1

- Contents on keywords

Case Study

- TF-IDF

Contents	User	TF	IDF	TF-IDF
C1	H1	0.0500	2.3	0.115
C2	H2	0.0375	2.3	0.086
C3	H3	0.0400	2.3	0.092
C4	H4	0.0600	2.3	0.138
C5	H5	0.0170	2.3	0.039

- $P_{i,m}$

User	Degree of centrality	$P_{i,m}$
H1	0.06	14.63
H2	0.12	29.26
H3	0.05	12.19
H4	0.02	4.87
H5	0.16	39.02

- $SR_{i,j,m}$

Contents	User	TF-IDF	$P_{i,m}$	$SR_{i,i,m}$
C1	H1	0.115	14.63	1.682
C2	H2	0.086	29.26	2.525
C3	H3	0.092	12.19	1.121
C4	H4	0.138	4.87	0.672
C5	H5	0.039	39.02	1.525

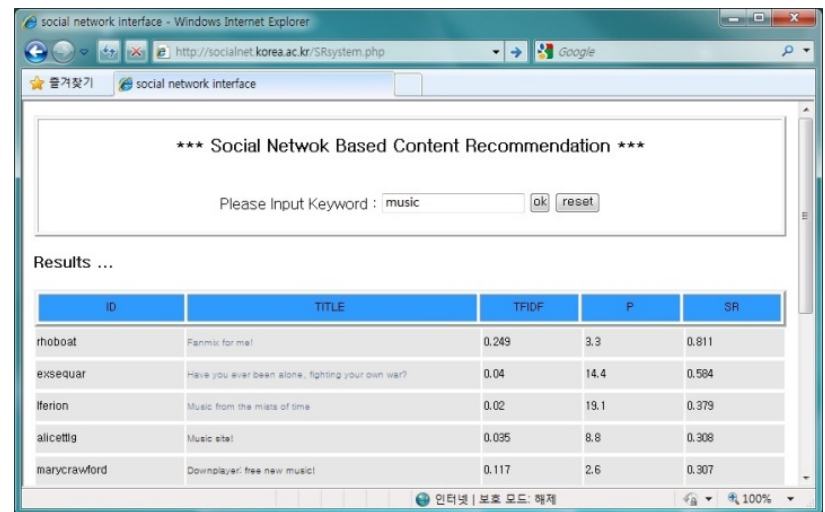
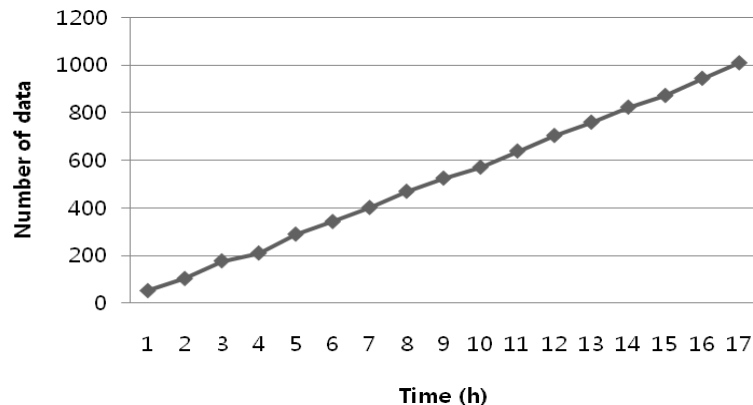
System Implementation and Evaluation

- Environments

- Data: MS Visual Studio 2008, MS SQL Server, C# language
- Web: PHP5, HTML, MS SQL

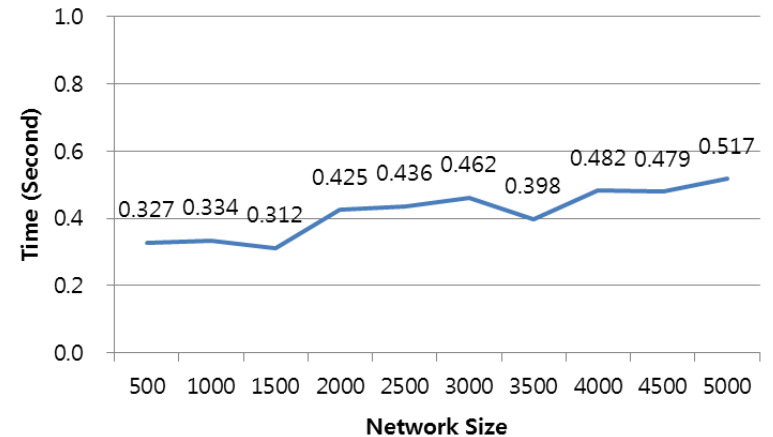
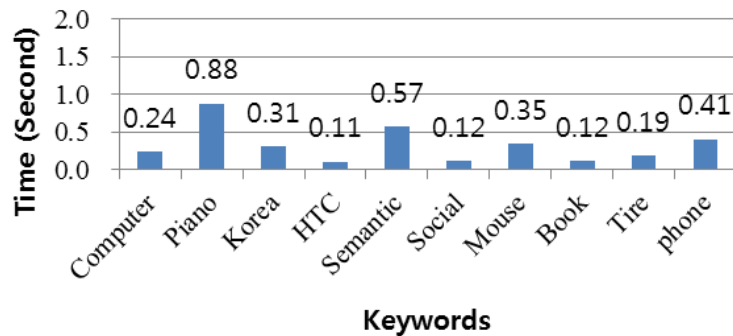
- Data

- 1,012 FOAF
 - 109,426 <foaf:knows>
- 12,373 RSS



System Implementation and Evaluation

- Response time



- Comparison

System	Recall	Precision	F1
Proposed	0.341	0.294	0.316
Pazzani and Billsus	0.245	0.215	0.231
Pan et al	0.289	0.318	0.302

Conclusion

- The method we proposed in this paper
 - Making recommendations of suitable content to users
 - Using TF-IDF and degree of centrality
- By using the proposed method
 - Users can actively share their experiences and knowledge
 - As well as collaboratively create and complement content
 - Effectively recommend content of high credibility to users
 - On rapidly developing social network services
 - Speed up content sharing
 - Among users
 - Based on the users' participation in social network services