Social Scope: Enabling Information Discovery on Social Content Sites

Authors: S.A.Yahia(Yahoo Inc) , L.V. Lakshmanan (UBC), Cong Yu (Yahoo Inc)

Presenter : Tony Wu

Date: 11/20/2014

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Overview

- Motivation
- Goals
- Architecture
- Social Content Graph Model
- Information Discovery Layer
- Collaborative Filtering Example
- Content Management Layer
- Information Representation Layer
- Conclusions

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Motivation

- Integration of social and content sites has led to the emergence of **social content sites** and **social content graphs**

- Traditional IR system only return results that are semantically relevant, but **not socially relevant**

	general (e.g., things to do)	categorical (e.g., family)	specific
with locations	32.36%	22.52%	8.37%
w/o locations	21.38%	5.34%	

Table 1: Summary Statistics of 10 Million Y!Travel Queries.

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Motivation - Example

- John has a day off from his conference in Toronto, he searches for “Toronto Attractions” on Y! Travel
- John has in the past visited a few “baseball fields” on Y! Travel and on Facebook he has many friends with interests in “baseball”
- Traditional IR systems would return all the generic popular attractions based on semantic relevance (e.g. CN Tower), but not baseball related (e.g. Rogers Center)
- Social-Scope solves this problem by incorporating social relevance its search results (e.g. consider John’s friends interests)

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Goals

- Query results must be both semantically relevant and socially relevant
- Efficient content management
- Effective navigation of search results

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Architecture

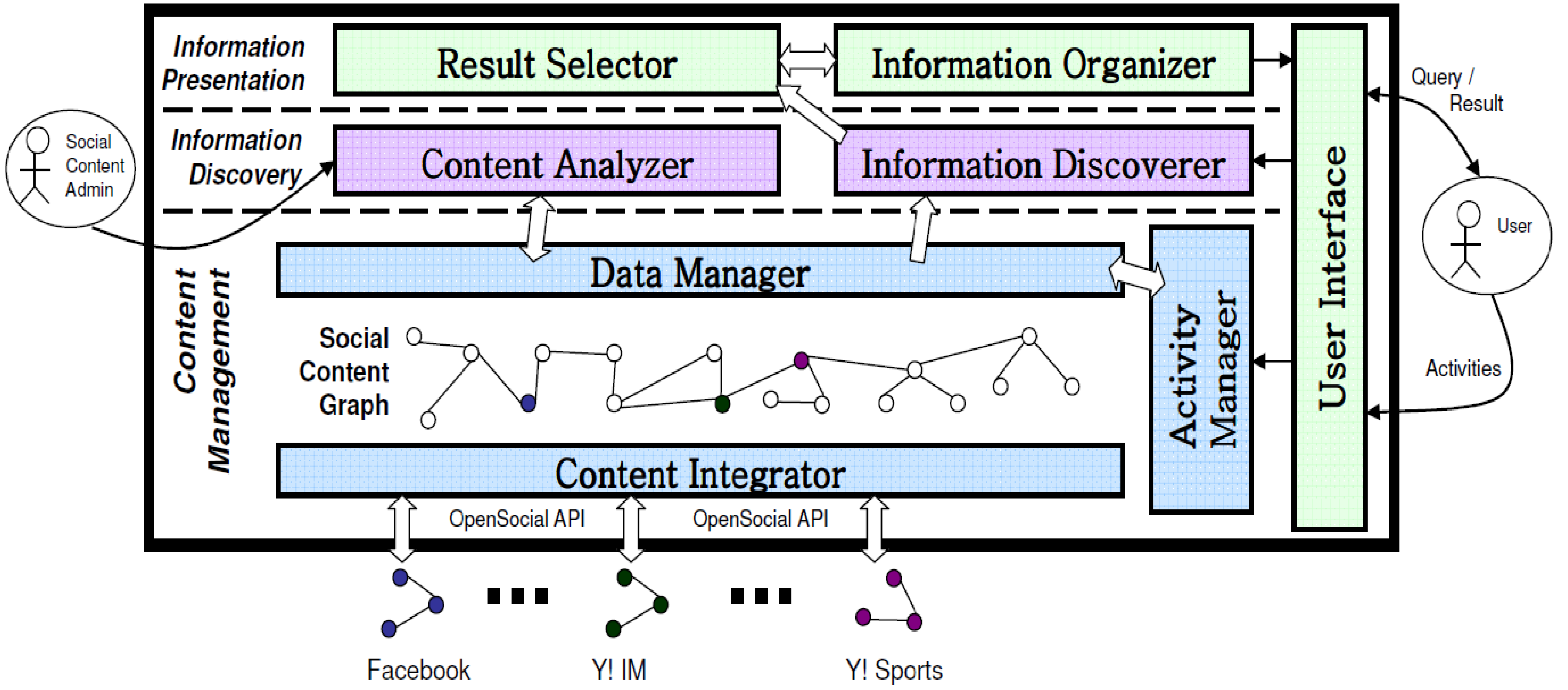

Figure 1: Architecture of SocialScope.

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Social Content Graph Model

- **Nodes**:
  - Represent users and entities like restaurants and attractions
  - Each node has an ID and a set of attributes (‘type’ attribute is mandatory)
  - E.g. n_1 = { id = 1; type = ‘user, traveler’; name = ‘John’}
- **Links**:
  - Represents relationships between nodes (e.g. “visited”)
  - Can also used to represent user tags
  - E.g. L_{12} (n_1, n_2) = {id = 12; type = ‘act, tag’; date = ‘2008-8-2’; tags = ‘Rockies baseball’}

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Information Discovery Layer

- Unary Operators:
  - **Node Selection**: $\sigma^{N}_{<C,S>}(G)$
  - **Link Selection**: $\sigma^{L}_{<C,S>}(G)$
  - Outputs and a set of nodes (for NS) or links (for LS) that satisfies condition C, and their scores based on the scoring function S

- Basic Binary Operators (Set theoretic Operators):
  - **Union** : $G1 \cup G2$
  - **Intersection** : $G1 \cap G2$
  - **Difference** : $G1/G2$
  - Applies to both links and nodes

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Information Discovery Layer

- Advanced Binary Operators:
  - **Composition:** $\boldsymbol{G = G_1 \odot_{<\delta\,,\,F>} G_2}$
    - Creates new links between the nodes in G1 and G2 satisfying the directional condition δ
    - Directional constraints have values src or tgt (e.g. δ = (src, tgt) means source node of G1 matches target node of G_2
    - *F* specifies the composition function of how the attributes of the new links should be determined based on the attributes of the input links
  - **Semi-Join:** $\boldsymbol{G_1 \ltimes_\delta G_2}$
    - Produces a sub-graph of G_1 where the links of G_1 matches the links of G_2 based on the directional condition δ
    - e.g. $G_1 \ltimes_{(tgt,src)} G_2$ produces a sub-graph of G_1 where the all links of the sub-graph have target nodes that matches the source nodes of G_2

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Information Discovery Layer

- Aggregate Operators:
  - **Node Aggregation:** $\gamma^{N}_{<C,d,att,A>}(G)$
    - Aggregates the links of the nodes based on aggregate function A satisfying condition C and directional parameter (src or tgt)
    - The aggregation function produces a new node attribute att
  - **Link Aggregation:** $\gamma^{L}_{<C,att,A>}(G)$
    - Replaces the set of links satisfying C and have the same source and target nodes where the new link has attribute att of which the value is determined by A

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Collaborative Filtering Example

Problem :

Recommendation of travel destinations to John based on his social network

Solution:

1. $G_1 = \ \sigma^{L}_{type='visit'} \left(G \ltimes_{src,src} \sigma^{N}_{id=101}(G)\right)$
   - User John and places he has visited
2. $G'_1 = \gamma^{N}_{type='visit',src,vst,A}(G_1)$
   - Set of destinations John has visited and stores in the vst att of node JOHN
3. $G_2 = \ \sigma^{L}_{type='visit'} \left(G \ltimes_{src,src} \sigma^{N}_{id\neq 101}(G)\right)$
   - Users other than John and the places they have visited
4. $G'_2 = \gamma^{N}_{type='visit', \, src, \, vst,A}(G_2)$
   - Set of destinations that other users have visited and stores in the vst att of the nodes

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Collaborative Filtering Example

Solution – Cont'd

5. $G_3 = G_1 \odot_{<\delta\,,\,F>} G_2$
   - $\delta = (tgt, tgt)$ and F is the composition function that computes the Jaccard similarity, which is stored in links produced from John to another user
6. $G_4 = \gamma^{L}_{<sim>0.5, type,\, A'>}(G_3)$
   - Replaces score of the links with similarity > 0.5 with string "Match"
7. $G_5 = \sigma^{L}_{type='visit'}\left(G \ltimes_{tgt,\, src} \sigma^{N}_{type='destination'}(G)\right)$
   - All destinations that users have visited
8. $G_6 = (G_4 \ltimes_{tgt,\, src} G_5) \odot_{<(tgt,src),\, sim_{sc}, F'>} (G_5 \ltimes_{src,tgt} G_4)$
   - For each of John's similarity network friends who has visited a destination, a new link is added from John to that destination. F' copies sim score from link of john to user to new link
9. $G_7 = \gamma^{L}_{C,\, sore,\, AVERAGE}(G_3)$
   - Replace set of links from John to destination node with attribute score, which is the average of the similarity score, score can be used to rank destinations

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Content Management Layer

- Provides **efficient storage** of indexes to support keyword-based queries
- **Inverted list (IL)** is used to store the item scores for each (tag, user) pair, each entry of the list consists of the form **(*i*, *score*$_k$(*i*,*u*))**, k = tag, I = item and u = user
- Score for each item is calculated by summing the item score across each IL of all (tag, user) pair
- **Problem**: Storing a list of scores for every (tag, user) pair can consumes a huge amount of storage since the number of items and users can be huge (requires 1 terabyte for a site with 100,000 users, million items and 1000 distinct tags!)

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Content Management Layer

- Clustering will reduce the storage complexity of storing inverted list
- Score of the cluster will be calculated as follows:

$$Score_k(i, C) = max_{u \in C}\, score_k(i, u)$$

- **Network-based Cluster:**

$$\frac{|\,network(u_1) \cap network(u_2)|}{|\,network(u_1)\ \cup network(u_2)|} \geq \theta$$

- **Behavior-based Cluster:**

$$\frac{|\,items(u_1) \cap items(u_2)|}{|\,items(u_1)\ \cup items(u_2)|} \geq \theta$$

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Information Presentation Layer

- Results are grouped to allow users to navigate the results more effectively

- **Social Grouping:**

$$\frac{|\,\boldsymbol{taggers}(\boldsymbol{i_1}) \cap \boldsymbol{taggers}(\boldsymbol{i_2})|}{|\boldsymbol{taggers}(\boldsymbol{i_1}) \cup \boldsymbol{taggers}(\boldsymbol{i_2})|} \geq \boldsymbol{\theta}$$

- Provides explanation to items to allow user realize **social provenance**:
  - Content-based strategy
  - Collaborative filtering strategy

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Conclusions

- Social Scope is able to search for results not only semantically relevant, but also social relevant
- Information discovery layer is based on a set of graph processing operators to return socially relevant results from the social content graph
- Content Management layer leverages clustering to efficiently store inverted index for key-word based search
- Search results are grouped in the presentation layer based on social grouping and descriptions are attached to results to realize social provenance

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.

Thank You!

Questions?

S. Amer-Yahia, L. Lakshmanan, and C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites . In Proc. Conf. on Innovative Data Systems Research (CIDR'09), January 2009.