A demo blog recommendation system for SmartScribo

Diana V. Zaiceva, Dmitry G. Korzun

Petrozavodsk State University
Department of Computer Science

This project is supported by grant KA179 of Karelia ENPI - joint program of the European Union, Russian Federation and the Republic of Finland

11\textsuperscript{th} FRUCT conference
April 23–27, Saint-Petersburg, Russia
Table of Contents

1 Multiblogging with SmartScribo
2 Blog recommendation service
3 Tags ranking model
4 Experiments
5 Conclusion
SmartScribo: Smart-M3 Application

- Distributed multi-agent architecture (clients, processors, mediators)
- Blogosphere is shared in the smart space
- Interaction with multiple blogs at many blog services
- Proactive blog retrieval, e.g., based on recommendation
- Personalization, e.g., user gets recommendation that is interested for him
Blogging

- Blog data
  - post starts discussion
  - discussion evolves with commentaries
  - tags describe post semantics

- Blogging context
  - post read&write: post status
  - given post $p$ and user $u$, $T_{pu}$ is the latest post access time
Blog Recommendation Service

Searching personally-interesting discussions in the blogosphere

- **Personal service**
  - track blog activity of user $u$: reading or sending post $p$
  - latest post access time $T_{pu}$

- **Proactive service**
  - user doesn’t make some additional actions

- **Read post**
- **Send post**
- **Notification about recommended posts**
- **Show recommended posts**
Basic Steps

1. User $u$ activity: post access time and other attributes

2. Tracking post status

3. Tag index $\text{data}(t, u)$ and rank $R_{tu}$

4. Blog discovery for top-ranked tags

5. Publishing blog recommendation
Architectural Design

Blog mediator

- subscribe to post status
- tags indexing and ranking
- query to blog search engine

Blog client

- subscribe to recommender
- proactively receive recommendations
- accept and enjoy
Simple Model

User $u$, post $p$, tag $t$

$n_p \geq 1$ is the sum number of tags in $p$

- $n_p$ is a post attribute kept in the blogosphere smart space

Blog mediator tracks status of $p$ and updates tags index

- $\delta_t$ is the time elapsed from the latest update (discrete: 0,1, 2, \ldots)

$R_{tu}$ shows importance of blogs with $t$ to $u$

$$R_{tu} = \begin{cases} R_0 / n_p, & \text{if } \delta_t = 0 \text{ (i.e., } t \text{ is new)} \\ (1 - \alpha) R_{tu}^{\text{old}} + \frac{\alpha \beta_0}{\delta_t n_p}, & \text{if } \delta_t \geq 1 \end{cases}$$

- $1 / n_p$ is topic-focus level of $p$
- $1 / \delta_t$ is freshness of $t$ (tag relevance)
- $0 < \alpha < 1$ is a tradeoff parameter
- $R_0$ - initial constant, $\beta_0$ constant for normalize time
Tag Relevance

Factor $1/\delta_t$ shows access activity for posts with $t$

The higher $\delta_t$ the less interest to $u$ in such posts

- Individual activity of $u$

$$\delta_t = \delta_t(u) = T_0 - \max_{t \in q, q \neq p} T_{qu}$$

- Collective activity: group $C$ influences $u$’s decisions

$$\delta_t = \delta_t(C) = T_0 - \max \left\{ \max_{t \in q, q \neq p} T_{qu}, \max_{v \in C \setminus \{u\}} \max_{t \in q} T_{qv} \right\}$$

where $T_0$ is the current time at blog mediator
Implementation

- **Blog Mediator:**
  - 4 modules on python with Smart-M3 Python KP Library
  - sqlite3 as database for ranks

- **Test KP Client:** console client, python, Smart-M3 Python KP Library
  - console client on python, Smart-M3 Python KP Library
  - file with test scenario

- **Blog Client:**
  - blog client on maemo5 platform,
  - implementation on PyQt, Smart-M3 Python KP Library
Test scenarios

- with 100% interest in topic
- with constant interest
- with variable interest
Running tests: 100% interest

- 30 reading posts with tag ‘dinner’
20 reading posts with tag 'dinner', 'president','biathlon'
'dinner' was about 65% cases, 'president' - 10%, 'biathlon' - 25%
recommendation is post with tag 'dinner'
Running tests: variable interest

- first 20 posts: ‘dinner’ was about 65% cases, ‘president’ - 10%, ‘biathlon’ - 25%
- next 10 posts: ‘dinner’ was about 25% cases, ‘president’ - 10%, ‘biathlon’ - 65%
- recommendation is post with tag ‘biathlon’
Proactive and personalized retrieval of blogs can be implemented on the Smart-M3 platform in computationally effective manner. Although the ranking model is simple, the approach allows discovering relevant blogs.

- SmartScribo project wiki (in Russian)
  http://oss.fruct.org/wiki/SmartScribo

- Open source
  http://gitorious.org/smart-scribo/smart-scribo

- Mailing list
  smart-scribo@cs.karelia.ru

Q&A