

Three Interesting Applications in Web, Email and Literature Management

Xuehua Shen

January 23, 2005

1 Web

A Personalized News Alerts. Right now, many people read news online routinely instead of buying a newspaper. Reading online, the user can save money(most online version of newspaper such as NYTimes.com has the same content as that of the paper version and at most require a free registration), save space at home (piled up newspaper may consume already crowded home space), selectively reading interesting articles and have multiple choices.

Generally the user not only read general news such as politics and business, but he has his possible multiple distinguished interests such as the news in his professional community and the news in his hobby. But at the same time, the user has to visit multiple web sites to browsing what is happening daily, weekly or monthly. If we have a provide a personalized News Alerts, a desktop software, which can automatically retrieve, rank and present the news of personal interest to the user, the user can save a lot time and do not miss important news.

Everyone can benefit from this software. The News articles (maybe in RSS format) and other web pages of World Web Wide will be crawled and filtered. For each person, we need only crawl a subset of web sites. This personalized News Alerts will have the filtering functionality (filtering uninteresting news), organization(organization by topic, date,etc), search (search by topic, date, etc) and mining (user can do comparative news reading to get different opinions about the same event.)

There is some challenges. First is how to model the user interesting. There are some clues about the user's personal interesting such as previous visited web sites and web pages. Second is to how to efficiently and robustly crawl multiple news web sites. Third is how to present the personalized news to the user. Fourth is how to provides some functionality to the user so that the user can manage the news, such as storing important news and annotating some news.

2 Email

Diary from Email. Some people have the habit to keep a journal (web blog, live journal, etc.). But sometimes, it is just a repeat of the email writing. On the other hand, the email provides a lot of information daily activities. Here we want to build a system to extract information events from the emails and compose a simple diary. Emails have a nice date associated with the contents.

Everyone can use this tool to automatically create a diary from the emails. This tool will provides the functionality of data extraction(entity extraction such as event and date) and composition of the diary.

There are a couple of challenges. First, how to efficiently extract event and date. Second, how to organize the event associated with the date.

3 Literature

Personal Literature Management Tool. Researchers store many papers on the local disk. Sometimes, it is hard to find the downloaded literature. So it is important to organize these papers and provide some functionality such as search to the user.

Every researcher will benefit from this tool. The personal literature data collection is the data this tool manages. The functionality will include search (find relevant papers) and classification. The user provides a hierarchy, then the system will associate each paper to several tags automatically. Of course, some papers will be tagged by the user so that some training data is provided.

The challenge of this project will be how to design and implement such a system and choose the best search and classification algorithms suitable for personal literature collection.