WOOF

Our Team:
Hadiwijaya, Alexander Daniel hadiwij2@illinois.edu
Jiang, Sharu sjiang19@illinois.edu
Luo, Hao haoluo3@illinois.edu
Nair, Ajay Jayan anair10@illinois.edu
Nguyen, Son nguyenb1@illinois.edu
Shi, Ruisheng rshi7@illinois.edu
Vijayakumar, Paul pvijaya2@illinois.edu
Yang, Fan fanyang5@illinois.edu
Zeng, Ruogu rzeng4@illinois.edu

Abstract:

We are implementing a news system that could crawl the news site, filter the news and classify the news into different categories. Then user could register and select their caring categories. The related news will be pushed to the user, we can then collect the users' feedback and modify the system accordingly.

Introduction:

If there is a satisfy button for all the news and feeds we see everyday.The satisfaction rate will be incredible low. Facebook are full of informations we are not interested. news website like CNN are redundant and overcomplicated. On the other side, there are some information we are really interested are not seen by us. We need a properly push system for the information. So, our ultimate purpose is to raise the satisfaction rate to 50+%.

Related Work:

Surprisingly, we only find one 'similar' product called stumbleupon. Stumbleupon sorted the information and list information according to user's interests. I think we there is a fundamental difference between us and stumbleupon is that stumbleupon is an amusing site. The information in stumbleupon is either updated or purely for amusing. We will focus on text and 'important' information, whatever important it is for user.

Methods:

From a user's perspective:
1. User will first go through a well sorted (graph) topics.

2. User will selected a ordered list(subraph) of topics.

3. User specify the source

4. User will identify some keywords/disgusting words for social medias. Also some special interests and blocks from specific contacts.

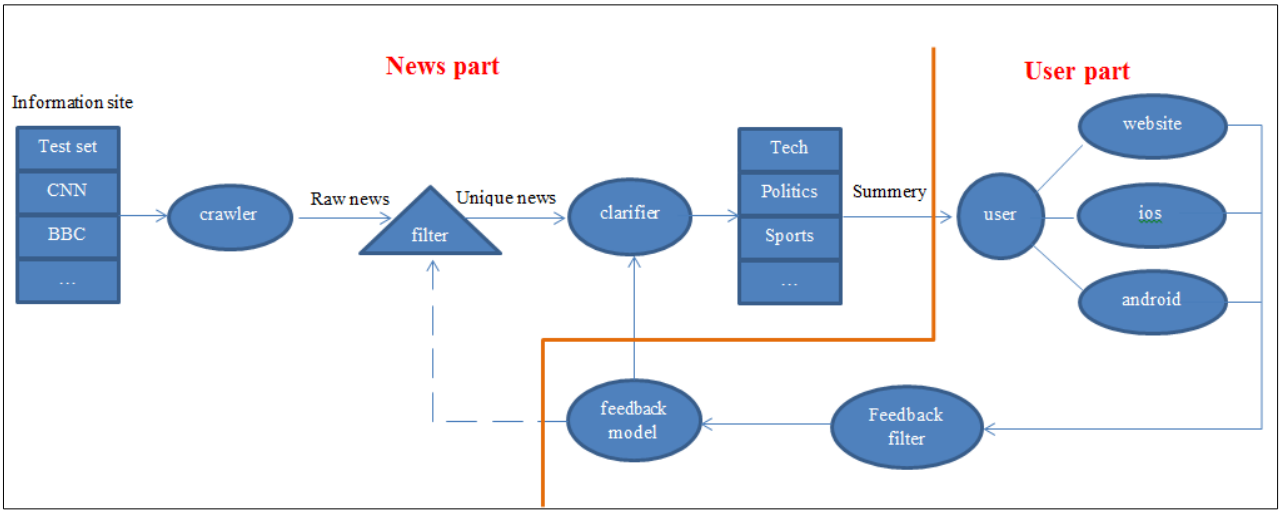
5. User will have a choice of local news/feeds first, or a section for local news etc.

6. User will receive an email full of interesting news and other information we crawled through website about the topic they selected. They will first login our website / phone ipad apps and access the links we provided.

7. User can logging to read the real time data.

We'll collecting feedback from:
Time of the viewing pages
Clicks through website
Clicks through apps
Commends and feedback: constructing commends and feedback subsystem. Customize the personal info demands

From developer's perspective



General speaking our system contains news part and user part.
From implementing step point of view, we have

Step 1:
website for internal use
crawler for news website like CNN

Step 2:
News filter implementation, throw the redundant and fake news.
Classified the news into different categories

Step 3:
user profile system
feedback system
(Keep expanding our crawler
Constructing the website for user)

Step 4:
Android apps
IOS apps

Initial approach:
1. From Limited category – full category

2. From test set of news – real time news – mining through the website

3. From english only- Chinese(Chinese has different source, they are incredible widely used) , French and Spanish if possible.

Future works:

profile sub systems -Identify the users with same interests,

Commends sub system

Localization sub system Localize the news feed, contain more local event for user to choose.

news-sub system Let user creating news themselves and public through our website

News sub-system for children

Appendix: Individual Responsibilities:

Team leader: Ruisheng Shi

Project manager: Fan Yang, Sharu Jiang, Qian Jia

Code reviewer: Paul Vijayakumar, Alexander Daniel Hadiwijaya, Ajay Jayan Nair, Son Nguyen

Model construction: Hao Luo, Sharu Jiang

Website construction: Ruogu Zeng, Alexander Daniel Hadiwijaya, Fan Yang, Ruisheng Shi

Crawler subteam: Ajay Jayan Nair, Paul Vijayakumar, Hao Luo, Sharu Jiang, Son Nguyen

References: list references of work related to your project or websites with related systems here.

[1] stumbleupon: http://www.stumbleupon.com/
[2] Jacobs, Paul S., and Lisa F. Rau. "SCISOR: Extracting information from on-line news." Communications of the ACM 33.11 (1990): 88-97.
[3] Lee, Chang-Shing, Zhi-Wei Jian, and Lin-Kai Huang. "A fuzzy ontology and its application to news summarization." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 35.5 (2005): 859-880.
[4] Dolan, Bill, Chris Quirk, and Chris Brockett. "Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.
[5] Kosala, Raymond, and Hendrik Blockeel. "Web mining research: A survey."ACM Sigkdd Explorations Newsletter 2.1 (2000): 1-15.
[6] Abel, Fabian, et al. "Analyzing user modeling on twitter for personalized news recommendations." User Modeling, Adaption and Personalization. Springer Berlin Heidelberg, 2011. 1-12.
[7] Faloutsos, Christos, “A Survey of Information Retrieval and Filtering Methods”
[8] Susan T. Dumais, “Personalized Information Delivery: An Analysis of Information Filtering Methods”, *Comminucations of the ACM*, *35*(12), 51-60, 1992