

WOOF

Team Bulldog:

Hadiwijaya, Alexander Daniel	hadiwij2@illinois.edu
Shi, Ruisheng	rshi7@illinois.edu
Yang, Fan	fanyang5@illinois.edu
Zeng, Ruogu	rzeng4@illinois.edu

Team Husky:

Jiang, Sharu	sjiang19@illinois.edu
Luo, Hao	haoluo3@illinois.edu
Nair, Ajay Jayan	anair10@illinois.edu
Nguyen, Son	nguyenb1@illinois.edu
Vijayakumar, Paul	pvijaya2@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. We have two subteams for this project.

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+%.

Methods:

From a user's perspective:

1. User will first go through a well sorted (graph) topics.
2. User will selected a ordered list(subgraph) 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:

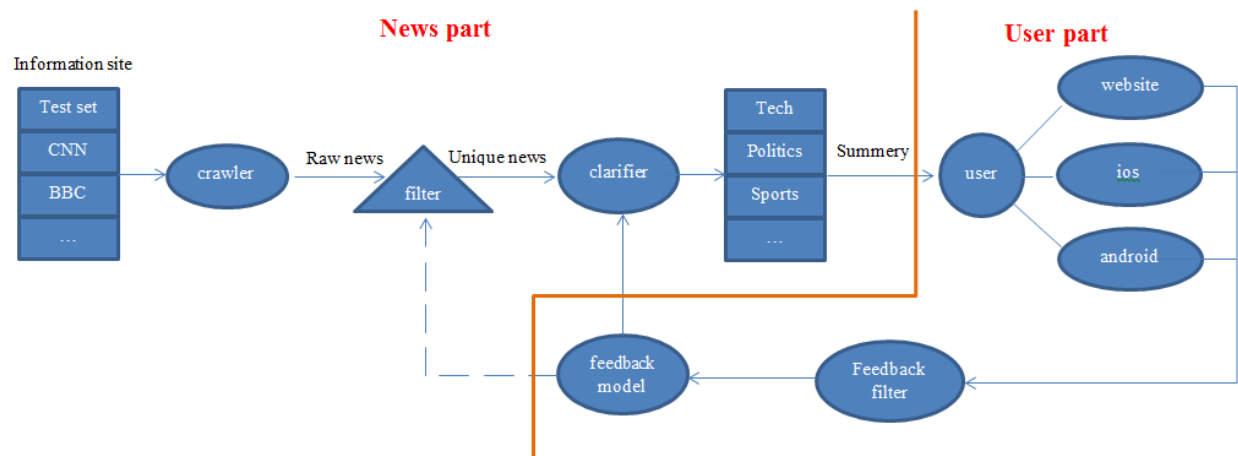
Time of the viewing pages

Clicks through website

Clicks through apps

Comments and feedback: constructing comments 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:

Bulldog: website for internal use

Husky: crawler for news website like CNN

Step 2:

Bulldog: News filter implementation, throw the redundant and fake news.

Husky: Classified the news into different categories

Step 3:

Husky: user profile system

Bulldog: feedback system

Husky: *Keep expanding our crawler*

Bulldog: *Constructing the website for user*

Step 4:

Husky: Android apps

Bulldog: 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 Bulldog

Team leader: Ruisheng Shi

Project manager: Fan Yang

Code reviewer: Alexander Daniel Hadiwijaya, Ruisheng Shi

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

Team Husky

Team leader: Ajay Jayan Nair

Project manager: Sharu Jiang

Code reviewer: Paul Vijayakumar, Ajay Jayan Nair, Son Nguyen

Model construction: Hao Luo, Sharu Jiang, Ruisheng Shi

News group members: Ajay Jayan Nair, Paul Vijayakumar, Hao Luo, Sharu Jiang, Son Nguyen

Related Work:

Stumbleupon

<http://www.stumbleupon.com/>

PROS: The idea of Stumbleupon is very similar to us.

CONS: Stumbleupon's system is linear and out of date. Also it is for amusement. We will focus on text and 'important' information, whatever important it is for user.

google news

<https://news.google.com/>

PROS: google news has millions of users, and they support customized preference. User can create label themselves

CONS: The preference is not really working well. After I select the only machine learning and world. The news are no significant different from the general news.

<http://www.newsisfree.com/>

PROS: this website actually provides an amazing amount of accurate news. The system also keep track of the news user has viewed.

CONS: You have to search it instead. It is not a push system. Users are lazy. They don't have an automatic summary system

<http://realtime.rediff.com/news>

A project that Ajay was closely involved in. Targets indian audience across the globe.

System is still imperfect though and lot of scope to improve. The idea though is still to push realtime news to the viewers.

<http://newzpile.com/>

PROS: they provides a crawling of lots of websites

CONS: they only keep a headline (title) of the news. Also, the UI is not user friendly.

<http://news.yahoo.com>

CONS: It has some categories on the side, but some if not most news can be found in other category.

<http://www.time.com/time/>

PROS: A bit more specific in categories, have some subcategories as well.

CONS: there are some unrelated post going to wrong category.

<http://feedly.com/index.html#explore%2Fartificial%20intelligence>

PROS: Categorized news search with option to follow feeds.

CONS: It is follow the website instead of the categories.

<http://www.douban.com/>

PROS: Choose the categories you are interested in, related articles/pictures will appear in the stream.

CONS: it is not news related. More like a categorized blogs

<http://www.brainpickings.org/>

PROS: all articles are selected by an expert reader. Choose the categories you are interested in, related articles will be pushed to your mailbox.

CONS: It is selected by human, limited by human's abilities.

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", *Communications of the ACM*, 35(12), 51-60, 1992