

‘LUPPEIT’

A LARGE RSS FEED RECOMMENDATION APPLICATION

by Faruk Kuşcan

Submitted to the Department of Computer Engineering
in partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Engineering

Boğaziçi University

January 2013

TABLE OF CONTENTS

ABSTRACT	3
1. PROJECT DESCRIPTION	4
1.1. PROBLEM DEFINITION	4
1.2. WHAT LUPPEIT OFFERS?	5
1.3. LUPPEIT'S BASICS	6
2. PROJECT REQUIREMENTS	7
2.1. FUNCTIONAL REQUIREMENTS	7
2.2. NON-FUNCTIONAL REQUIREMENTS	8
3. HOW LUPPEIT WORKS?	9
4. SCREENS	10
5. SYSTEM INFORMATION	17
5.1. SYSTEM DESIGN	17
5.2. USED TECHNOLOGIES	18
5.3. DATABASE	19
5.4. RECOMMENDATION	19
6. DEVELOPMENT SPECIFICATION	20
6.1. USED SOFTWARE DEVELOPMENT TOOLS	20
7. CONCLUSION & FUTURE WORK	21
APPENDIX	21

ABSTRACT

Project Name : LuppeIt

Project Team : Faruk Kuscan

Term : 2012/13 I.Semester

Keywords : Social Media Analysis, Recommendation, RSS

Summary : LuppeIt is a tool for any kind of user who wants to find links in internet which he/she would be interested in more easily. Also it is a good social media analysis tool. LuppeIt is a web project which a user can be registered. For registered user, LuppeIt keeps the track of all user actions in the application. According to this data, LuppeIt recommends user related links coming from RSS feeds.

1. PROJECT DESCRIPTION

1.1 PROBLEM DEFINITION

The internet has lots of news, shares, blog posts and many other different kinds of information in it. However it is also a very big sea so sometimes we can not find what we really need in the internet. When someone needs to find something in internet, search engines are used like Google. This is a way of searching for something in internet when we “need” that specific information.

Also there is another kind of search. That is when we do not need that specific information. For example, we type nytimes.com and see what news are posted today. However, there are hundreds of news posted each day. Most of those news are not interesting for all users of nytimes.com. In this part, the other kind of search starts. That is a manual search for that specific user to find content which he or she will be interested in to read.

This is not a good way to find the content what the user is interested in. In the first part, the user should filter all the web by opening a web page. The web sites are not many. If the user doesn't know abcd.com, he or she will not be able to see the content he or she may be interested in. Also the second part, the user should search manually by clicking on categories etc.

The problem here has two main points. First one is the following: Users can not know all the web sites in the internet where the content they are interested in are published. Second point is the manual search made by user to find the content they are interested in. Their interests are surrounded by many other content which they do not want to read.

1.2. WHAT LUPPEIT OFFERS?

LuppeIt offers a solution to both of the problems analyzed in the first part of this paper. LuppeIt is a web application which gathers content from all around the web for its users and filters these content according to its users' interests. LuppeIt does this filtering job by given data from the user and all the actions made by that user on LuppeIt.

As the problem is analyzed in the first part, LuppeIt finds solution to the problem as told in the following.

The problem has two major points. First one was that the user does not know all the web pages. LuppeIt gathers its content from a large number of RSS feeds. Number of RSS feeds which LuppeIt is integrated to will increase each day. In this way, LuppeIt will be able to offer a very big amount of content to its users.

Second problem was the manual search done by users of internet to find the content they would be interested in. LuppeIt offers this problem a solution. LuppeIt learns and knows about its users. It keeps the track of each user individually and according to the data like tags, categories, content etc. LuppeIt learns more about its users. When a user comes to LuppeIt and logs in, LuppeIt generates a result set of feed entries from all around the web according to that individual user's past actions on LuppeIt. These actions are like, dislike, view a share, tag a share, approve a tag, disprove a tag etc. However, LuppeIt is developed in a fashion that any kind of new actions can be easily implemented into working state of the code. This generic action structure is loggable and any kind of actions' logs can be used in the recommendation algorithm which generates the result set of news, shares etc. that individual user will see when he or she logs in to the system.

1.3.LUPPEIT'S BASICS?

Share

Share is either a url coming from an RSS feed that LuppeIt uses as resource.

Category

In LuppeIt all shares are seen under a category like “World News, Politics, etc.”

These categories can not be defined by users. A share can not be related to multiple categories.

Tag

Relevant words about a share. A share can have multiple tags. Users can add new tags. Tag has a truth value in LuppeIt for finding better tags for shares.

Resource

The source web page of a share is resource. As an example, if a share's url is <http://www.ntvmsnbc.com/id/12345>, the resource is “ntvmsnbc.com”.

Like/Dislike (Luppe/Bury)

Users will be able to like or dislike a share. This will help LuppeIt to learn more about the user and more about the average of all users and resources.

Interest

Users will be able to mark some categories as their interests.

These are basic definitions that we need to know to understand LuppeIt.

User Types

There are two types of users on Luppelt: Anonymous user, registered user. For registered users, Luppelt will look to their user past and behave according to that. However, for anonymous user, Luppelt will act according to the general average of all users likes and dislikes.

2.REQUIEMENTS

2.1.FUNCTIONAL REQUIREMENTS

LOGIN / REGISTRATION

Users will be able to login to the system. The whole system will work according to user data when a user is logged in. If the user is not logged in an anonymous user will be using the system. However, if the user wants to find his interests from the web, he or she must register and log in.

VIEW / LUPPE / BURY A SHARE

A user will be able to view a category. In that category page, user relevant content of that category will be gathered. The user will be able to click on a share in that category and view that link. Also, the user will be luppe (like) or bury(dislike) those shares. View, luppe and bury actions will be stored by Luppelt to be used to know the user better. Result sets created for category contents for user etc. will be generated according to these data.

CHANGE INTERESTS

A user will be able to tell his or her interests during registration. Also after the registration he or she will be able to change his or her interests.

TAG A SHARE

A registered user must be able to add a tag to a share.

APPROVE/DISPROVE A TAG

A registered user must be able to agree or disagree with a tag given to a share. With a very simple user interface, users must be able to do this to find which tag is more about a share.

2.2.NON-FUNCTIONAL REQUIREMENTS

- Availability
 - The application will be available on web interface.
 - The application should run 24/7.
 - Users from all around the world must be able to use it in English.
- Performance
 - Recommendations must be gathered and exposed to each user in a short time.
 - Cache levels must be used to fulfill this requirement.
 - User past may be cached in each login of a user.
- Security
 - Users' login data must be stored in a secure way.
 - Users' demographic data must be stored in a secure way.
 - Tag mechanism must be implemented in a way that users can add tags to shares etc. However, tags must be securely controlled to avoid misleading tags.
- Usability
 - LuppeIt must have an easy to use user interface.
 - Luppe and Bury actions must be analyzed in a page to the users.

- Technical
 - Play! Framework will be used in implementation. (An MVC framework works with Java)
 - MySQL database will be used as data storage.
 - Groovy language will be used in generating the views with Play! Framework's template engine
 - ROME API will be used to integrate to RSS feeds and crawl content of the RSS into LuppeIt.

3.HOW LUPPEIT WORKS?

Working Environment

LuppeIt is being implemented with Java. An MVC framework called Play! is being used. This MVC framework is connected to a MySQL database. The application is deployed on Play! Framework's built-in application server. The current version of LuppeIt is deployed on a live server.

<http://luppeit.com>

<http://swe.cmpe.boun.edu.tr:7880>

4. SCREENS

Unregistered user top news page

Luppelt

[Login](#) [Registration](#)

Top News

Most Recent

Categories

Troubles Continue For 787 Dreamliner With Groundings In Japan, U.S.

21 hours ago

An All Nippon Airways (ANA) plane was forced to make an emergency landing on Wednesday because of an apparent electrical failure. ANA and Japan Airlines have grounded all their 787s in response. Later Wednesday, the FAA grounded the U.S. fleet, which effectively grounds all 787s.

Business & Investing | [npr.org](#)

1 0

You Can Use Facebook for Free Cellphone Calls Now

20 hours ago

Wait, what?

Social Media | [yahoo.com](#)

1 0

Facebook unveils 'Graph Search,' a search engine for Facebook

20 hours ago

Facebook (FB) on Tuesday held a press conference at its Menlo Park, California headquarters where it unveiled "Graph Search," a search engine of sorts designed to allow users to search for content on Facebook. According to Facebook CEO Mark Zuckerberg, Graph Search allows users to search for news posts, status updates, photos, locations and more within the Facebook Social

Categories

Arts & Culture

Automotive

Business & Investing

Fashion

Health

Movies

Music

Politics

Science

Social Media

Sports

Technology

Travel

Web

World News

All Categories

Unregistered user most reconds page

Luppelt

[Login](#) [Registration](#)


Top News

Most Recent

Categories

TSX ends broadly higher; financials, U.S. data lifts

59 seconds ago



TORONTO (Reuters) - Canada's main stock index finished higher on Thursday in a broad rally led by financial and resource stocks and spurred in part by encouraging U.S. economic data and corporate results. The Toronto Stock Exchange's S&P/TSX composite index climbed 65.91 points, or 0.52 percent to 12,674.73. All 10 of the index's main sectors advanced. (Reporting by Solarina Ho)

Business & Investing | [yahoo.com](#)

0 0 0

Oil hits 4-month high on US economic data, Algeria

59 seconds ago

NEW YORK (AP) — The price of oil rose to a four-month high Thursday on positive reports about the U.S. economy.

Business & Investing | [yahoo.com](#)

0 0 0

Why Google Isn't Scared of Facebook's Graph Search

Categories

Arts & Culture

Automotive

Business & Investing

Fashion

Health

Movies

Music

Politics

Science

Social Media

Sports

Technology

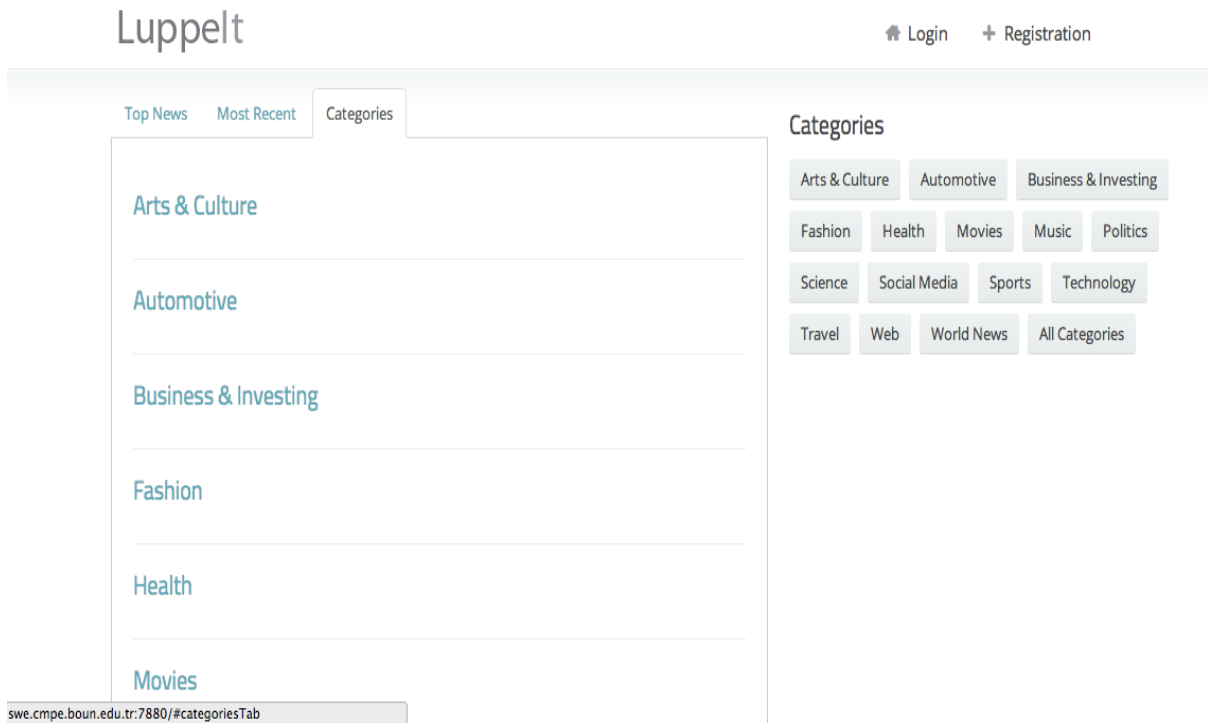
Travel

Web

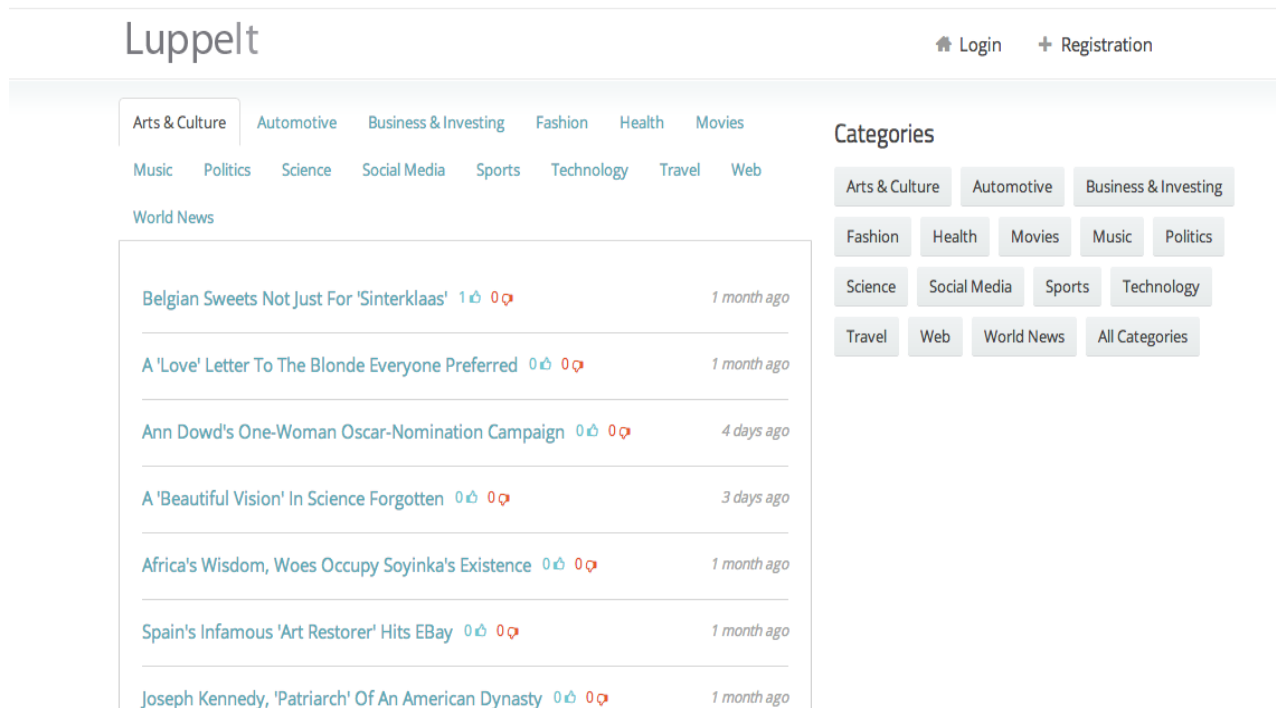
World News

All Categories

Unregistered user landing categories





Unregistered user categories page



Unregistered user single category page

Luppelt



 Login  Registration

Arts & Culture

Top News



Most Recent

Belgian Sweets Not Just For 'Sinterklaas'

1  0 



1 month ago

A 'Love' Letter To The Blonde Everyone Preferred

0  0 



1 month ago

Ann Dowd's One-Woman Oscar-Nomination Campaign

0  0 



4 days ago

A 'Beautiful Vision' In Science Forgotten

0  0 



3 days ago

Africa's Wisdom, Woes Occupy Soyinka's Existence

0  0 

1 month ago

Spain's Infamous 'Art Restorer' Hits EBay

0  0 

1 month ago

Categories

Arts & CultureAutomotiveBusiness & Investing


FashionHealthMoviesMusicPolitics

ScienceSocial MediaSportsTechnology

TravelWebWorld NewsAll Categories

Registered user top news page

Luppelt

 faruk.kuscan@gmail.com

Top News

Most Recent



Your Categories

Troubles Continue For 787 Dreamliner With Groundings In Japan, U.S.

21 hours ago

An All Nippon Airways (ANA) plane was forced to make an emergency landing on Wednesday because of an apparent electrical failure. ANA and Japan Airlines have grounded all their 787s in response. Later Wednesday, the FAA grounded the U.S. fleet, which effectively grounds all 787s.

Business & Investing | [npr.org](#)



1  0 

Troubles Continue For 787 Dreamliner With Groundings In Japan

22 hours ago

Troubles continue for Boeing's new 787 Dreamliner. An All Nippon Airways (ANA) plane was forced to make an emergency landing on Wednesday because of an apparent electrical failure. ANA and Japan Airlines have grounded all their 787s in response. That means nearly half the global fleet of Boeing's flagship plane is now sitting on the ground.

Business & Investing | [npr.org](#)

0  0 

Facebook unveils 'Graph Search,' a search engine for Facebook

20 hours ago

Facebook (FB) on Tuesday held a press conference at its Menlo Park, California headquarters

Waiting for 22 minutes ago

Categories

Arts & CultureBusiness & InvestingMoviesMusic

PoliticsScienceSocial MediaTravel

World NewsAll Your Categories

Registered user most recent page

Luppelt

faruk.kuscan@gmail.com


Top News

Most Recent

Your Categories

TSX ends broadly higher; financials, U.S. data lifts

4 minutes ago



TORONTO (Reuters) - Canada's main stock index finished higher on Thursday in a broad rally led by financial and resource stocks and spurred in part by encouraging U.S. economic data and corporate results. The Toronto Stock Exchange's S&P/TSX composite index climbed 65.91 points, or 0.52 percent to 12,674.73. All 10 of the index's main sectors advanced. (Reporting by Solarina Ho)

Business & Investing | yahoo.com

Oil hits 4-month high on US economic data, Algeria

4 minutes ago

NEW YORK (AP) — The price of oil rose to a four-month high Thursday on positive reports about the U.S. economy.

Business & Investing | yahoo.com

Why Google Isn't Scared of Facebook's Graph Search

Categories

Arts & Culture

Business & Investing

Movies

Music

Politics

Science

Social Media

Travel

World News

All Your Categories

Registered user interests list page

Luppelt

faruk.kuscan@gmail.com

Top News

Most Recent

Your Categories

Arts & Culture

Business & Investing

Movies

Music

Politics

Science

Categories

Arts & Culture

Business & Investing

Movies

Music

Politics

Science

Social Media

Travel

World News

All Your Categories

Registered user all interests page

Luppelt

faruk.kuscan@gmail.com

Arts & Culture

Business & Investing

Movies

Music

Politics

Science

Social Media

Travel

World News

Categories

Arts & Culture

Business & Investing

Movies

Music

Politics

Science

Social Media

Travel

World News

All Your Categories

A Cooler Roof For A New 'Cat' 0 0 12 hours ago

Chicken With Gatorade: The Oddball Pleasures Of '... 0 0 4 hours ago

New World, Old Evil In Tracy Chevalier's 'Runaway' 0 0 9 hours ago

Belgian Sweets Not Just For 'Sinterklaas' 1 0 1 month ago

How A 'Madwoman' Upended A Literary Boys Club 0 0 2 hours ago

Oprah Interview Adds To Armstrong's Complicated M... 0 0 22 hours ago

Rereading The Classics: Lessons Learned The Secon... 0 0 51 minutes ago

Registered user single interest's shares page

Luppelt

faruk.kuscan@gmail.com

Movies

Top News

Most Recent

Watch The Golden Globes With Us, Where The Drinks... 0 0 4 days ago

Ann Dowd's One-Woman Oscar-Nomination Campaign 0 0 4 days ago

Tom Hooper On The Magic Of 'Les Miserables' 0 0 26 days ago

Blame The Movies As We Mourn? Or Turn To Them For... 0 0 26 days ago

Last-Minute Box Sets: Home Video For Everyone On ... 0 0 26 days ago

Reliving An 'Impossible' Catastrophe 0 0 26 days ago

Categories

Arts & Culture

Business & Investing

Movies

Music

Politics

Science

Social Media

Travel

World News

All Your Categories

Registered user view share page

Luppelt

faruk.kuscan@gmail.com

Troubles Continue For 787 Dreamliner With Groundings In Japan, U.S.

npr.org

21 hours ago

An All Nippon Airways (ANA) plane was forced to make an emergency landing on Wednesday because of an apparent electrical failure. ANA and Japan Airlines have grounded all their 787s in response. Later Wednesday, the FAA grounded the U.S. fleet, which effectively grounds all 787s.

Read More >>

CATEGORY

Business & Investing

TAGS

+ Add Tag

japan ✓ X

Lupped

Bury

Registered user view original share page

Luppelt

faruk.kuscan@gmail.com

New User? Register | Sign In | Help

Make Y! My Homepage

🔔 | 📧 Mail | My Y! | 🏠 Yahoo!

YAHOO! NEWS

Q Search

Search Web

HOME | U.S. | WORLD | BUSINESS | ENTERTAINMENT | SPORTS | TECH | POLITICS | SCIENCE | HEALTH | BLOGS | LOCAL | POPULAR

Q News Search

News Search

Videos | Photos | Katie's Take | Trending Now | This Could Be Big | Weather | Who Knew? | Newsmakers | Enterprise

DISCOVER YAHOO! WITH YOUR FRIENDS

📘 Login

Facebook unveils 'Graph Search,' a search engine for Facebook

By Zach Epstein | BGR News – Tue, Jan 15, 2013

Email

📘 Share

🐦 Tweet

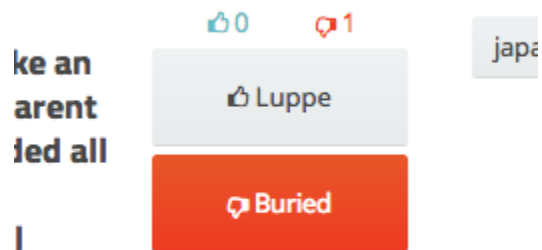
📘 Share

🖨 Print

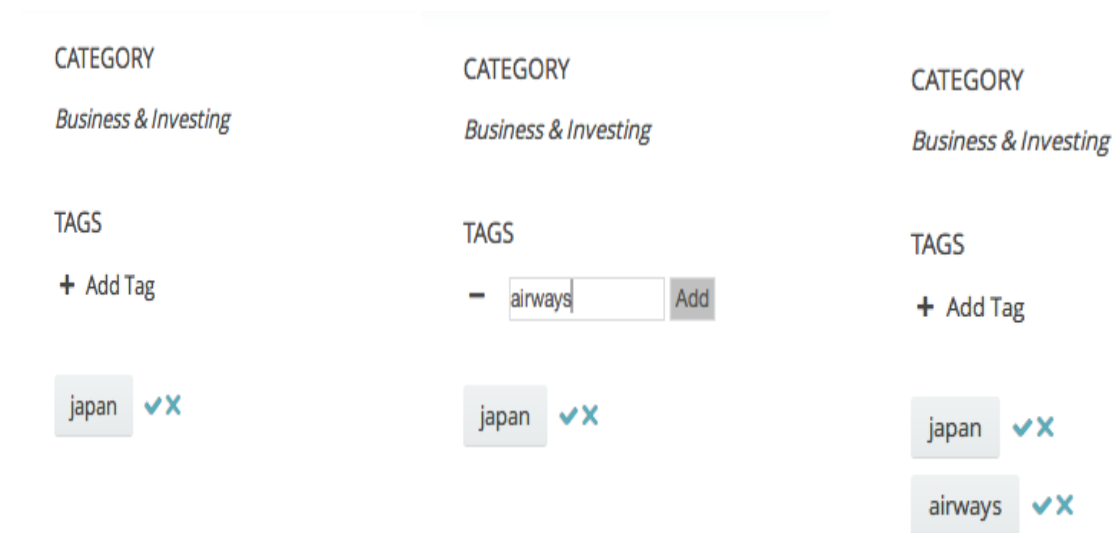
NEWS FOR YOU »

- Team USA announces roster for World Baseball Classic
- Khloe Kardashian Addresses Tabloid Cover Stories On Marriage, Paternity: They're 'Untrue,' 'Unfair'
- Subway footlong sandwich really 11 inches?
- Dad Uses Vacuum to Create Daughter's Perfect Ponytail
- 'Argo' Actor Comes Out
- Soccer-Suarez admits dive but says is picked on by media
- Minaj, Carey fail to boost "American Idol"

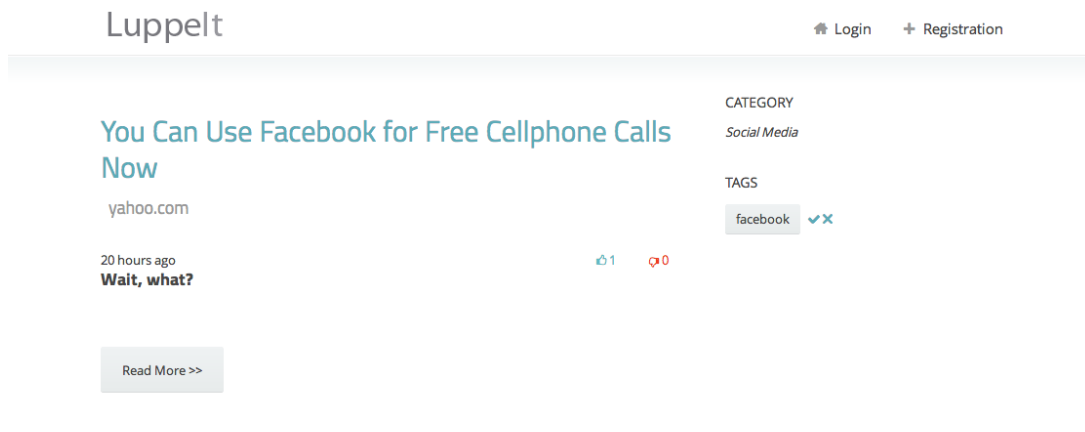
Registered user luppe/bury share



Registered user add tag to share



Unregistered user single share page



5. SYSTEM INFORMATION

5.1. SYSTEM DESIGN

Luppelt is implemented with Java programming language. A web MVC framework which has a very rapidly growing community name Play! Framework is used to implement this application.

There are five main parts of the application.

1. Controllers

These are controller objects which implements the logic behind each screen of the application. Requests are caught in servlet wrapper of Play! Framework and application related logic is handled in these controllers. Controllers are the bridge between models (indirectly DB) and views (indirectly users).

2. Database

There is a single database in Luppelt which stores the data of shares, tags, users, categories etc.

3. Models

These are classes which represent tables of DB in code. For example there is a User.java class which represents a single record in “user” table in the DB.

4. Views

These are the user interfaces being generated by controllers and exposed to users via browsers. Play! Framework’s Groovy template engine is used to generate views in LuppeIt.

5. Job

There is a single job class which runs periodically on the background of application. This job runs on application start and every five minutes after that. This job looks for all rss resources next feed date fields. This job gets data from RSS feeds if next feed date of an RSS is in past while the job is running.

There are 3 main threads in the application which handles requests. Also there is a connection pool of application for DB connections. Minimum size of this pool is 5 and maximum size is 30. This is an important setting done in LuppeIt which affects DB access time for users.

ACTION STRUCTURE

A generic action structure is built in LuppeIt. In the current working state of implementation, there are actions like viewing, lopping, burying a share, adding a tag to a share, approving, disproving a share etc. However, the generic structure of “Action” interface in LuppeIt lets to create new types of actions and call those actions wherever we want and log these actions. The recommendation algorithm in LuppeIt also can be merged with any kind of actions - which implements (extends) the generic action interface- logs. The thing I am trying to analyze is

that with small changes in code of LuppeIt, we can increase, decrease priority of lopping a share, burying a share, viewing a share while recommendation process. Also we can add new actions very easily, just by creating a new class which implements generic ActionEndPointIF.java interface.

5.2. USED TECHNOLOGIES

Play! Framework: MVC structure

MySQL: Database needs

Groovy Template Engine: Generating views

Rome API: Integrating with RSS feeds

Git: For versioning needs

GitHub: For remote repository needs

SmartGit: As Git Client

Navicat for MySQL: For database user interface

5.3. DATABASE

ER diagram of database can be found in the following URL:

http://github.com/kuscan/luppeit/blob/master/LuppeIt_ER_Diagram.pdf

5.4. RECOMMENDATION

Recommendation mechanism in LuppeIt is working in the following fashion. There is a distinction at the start. If the user is registered, LuppeIt acts different. There is always a UserPast object in the cache. There is a single UserPast object in cache for each registered and logged in user. Also there is one other single object which can be used by all anonymous users online. These UserPast object has all user past data (shares that he/she lopped, tags of shares which are

lupped/buried by user, viewed shares, viewed share's tags, viewed share's resources etc.). Also each of these parameters has a coefficient (statically defined in the code). With respect to the coefficient of these parameters, all shares inserted into DB in last 7 days gets a rank from 0 to infinity. The highest rank becomes first in the result set.

If user is registered, these recommendation process is done only for categories which are in that user's interests list.

In the recommendation process, lastly, the headline of the RSS feed entry is used. All the words in the headlines are lowercased and all the punctuations in these words are deleted. Then a special string is appended before each word. Each of these strings are inserted into DB as tags of that share (however, these shares are not seen in user interface). In this way, the more words a share has common with another share relates to each other. If a user makes a positive action on LuppeIt on a share, shares with similar headlines are detected and listed with higher priority.

6. DEVELOPMENT SPECIFICATION

6.1. USED SOFTWARE DEVELOPMENT TOOLS

Git versioning tool is used for versioning.

GitHub is used as a remote repository.

<http://github.com/kuscan/luppeit/>

SmartGit is used as git client.

7. CONCLUSION & FUTURE WORK

LuppeIt is a tool for users who are looking for content they would be interested in. This application tries to make a difference with solutions it offers to the two main problems described in the first part of this report.

As a future work the following improvements can be done:

1. Social media analysis for making better recommendation

For now, LuppeIt only recommends shares according to the data a user logs into DB via his/her actions on LuppeIt. However, with social media analysis, relations between different users can be created and better result sets can be obtained from recommendation.

2. A mobile user interface can be done. Users are able to read news etc. in any place via smart phones. A user friendly mobile application can be done as a second medium other than web.

APPENDIX

The video of LuppeIt:

<http://www.youtube.com/watch?v=seQPpdGr26Q>