

GUJARAT TECHNOLOGICAL UNIVERSITY

Chandkheda, Ahmedabad



Lalbhai Dalpatbhai College of Engineering

Department of Information and Technology Engineering

A report on

PURE TRENDING CONTENT

B. E. SEM – VII

Submitted by,

Kalpit Akhawat 150283116001

Jigar Bhatt 150283116002

Jimish Fotariya 150283116008

Hardik Solanki 150283116026

Academic year – 2017

Faculty guide,

Manoj Patel – Assistant Professor

Head of department,

Hiteishi Diwanji – Associate Professor

Index

Introduction	01
I. Problem summary	01
II. Scope	02
III. Aim and objectives	03
IV. Problem specifications	04
V. Brief literature review and prior art search (PAS)	05
VI. Technology and tools required	06
Design	09
I. Functional requirements	09
II. Non-functional requirements	10
III. Data flow diagram	11
IV. UML diagrams	15
V. Explanation of work	19
Implementation	29
I. Performance	29
II. Design for ergonomics	32
Summary of the results	36
I. Problem summary	36
References	
Appendix	
Conclusion and flow	

Acknowledgement

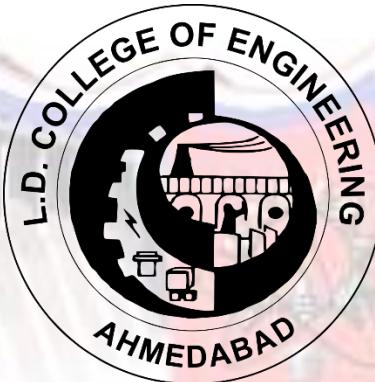
We consider it a privilege to be associated with the L.D. College of Engineering & Technology, Ahmedabad in this academic endeavor. We express our heartfelt thanks to our Guide Mr. Manoj Patel, Assistant Professor in Department of Information and Technology Engineering, for his valuable guidance, continued interest throughout the course of this work and encouragement towards the successful completion of this project. We are thankful to Mrs. Hiteishi Diwanji, Head of department in Department of Information and Technology Engineering for providing necessary arrangements for project works. We are also very much thankful to all the faculty members for their valuable suggestions and comments during our project work. We are thankful to all the people who joined as part of making this journey of fulfilling of this Project. We are grateful to Gujarat Technological University for giving us a wonderful new subject for improving our knowledge and research capabilities.

Certificate

Lalbhai Dalpatbhai College of Engineering

Department of Information and Technology Engineering

Academic year – 2017



Date, 2017

This is certified that the project entitled Pure Trending Content is submitted by Kalpit Akhawat (150283116001) under the guidance of below mentioned Faculty and the Head of Department for VII semester in the four walls of L.D. College of Engineering during the academic year.

Faculty guide,

Manoj Patel – Assistant Professor

Head of department,

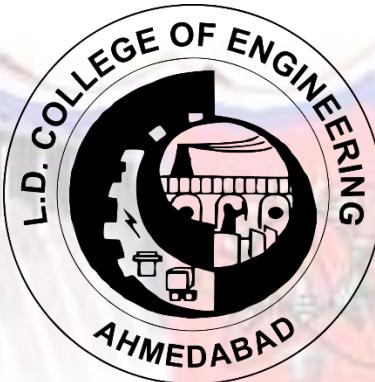
Hiteishi Diwanji - Associate Professor

Certificate

Lalbhai Dalpatbhai College of Engineering

Department of Information and Technology Engineering

Academic year – 2017



Date, 2017

This is certified that the project entitled Pure Trending Content is submitted by Jigar Bhatt (150283116002) under the guidance of below mentioned Faculty and the Head of Department for VII semester in the four walls of L.D. College of Engineering during the academic year.

Faculty guide,

Manoj Patel – Assistant Professor

Head of department,

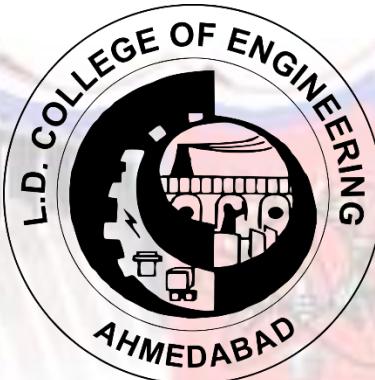
Hiteishi Diwanji - Associate Professor

Certificate

Lalbhai Dalpatbhai College of Engineering

Department of Information and Technology Engineering

Academic year – 2017



Date, 2017

This is certified that the project entitled Pure Trending Content is submitted by Jimish Fotariya (150283116008) under the guidance of below mentioned Faculty and the Head of Department for VII semester in the four walls of L.D. College of Engineering during the academic year.

Faculty guide,

Manoj Patel – Assistant Professor

Head of department,

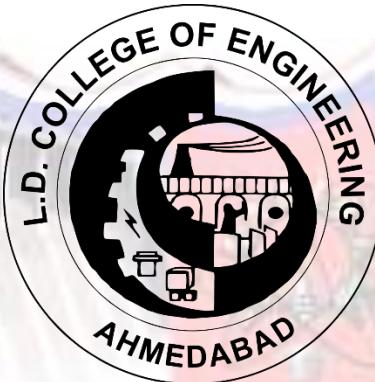
Hiteishi Diwanji - Associate Professor

Certificate

Lalbhai Dalpatbhai College of Engineering

Department of Information and Technology Engineering

Academic year – 2017



Date, 2017

This is certified that the project entitled Pure Trending Content is submitted by Hardik Solanki (150283116026) under the guidance of below mentioned Faculty and the Head of Department for VII semester in the four walls of L.D. College of Engineering during the academic year.

Faculty guide,

Manoj Patel – Assistant Professor

Head of department,

Hiteishi Diwanji - Associate Professor

PLAGIARISMA

79% Unique

Total 62150 chars (**2000 limit exceeded**) , 326 words, 19 unique sentence(s).

Essay Writing Service - Paper writing service you can trust. Your assignment is our priority! Papers ready in 3 hours! Proficient writing: top academic writers at your service 24/7! Receive a premium level paper!

STORE YOUR DOCUMENTS IN THE CLOUD - 1GB of private storage for free on our new file hosting!

Results	Query	Domains (original links)
Unique	All of which is to say that there will probably be a wide range	-
16,900,000 results	Easy, no	realsimple.com cookstr.com thehealthyfoodie.com profiles.nlm.nih.gov thedailymeal.com cheesecake.com Free Download Mozilla Firefox® Web Browser www.mozilla.org Download Firefox - the faster, smarter, easier way to browse the web and all of Yahoo 12345Next16,900,000 results
Unique	What is TIMESWEN	-
Unique	TIMESWEN is a web portal and related collection of Internet things & services	-
Unique	No employing editors to repurpose existing content from partners at popular organizations	-
Unique	So, it should be to Improve trending	-
Unique	Also, rich content after that from another resource & check purity of their contents	-
Unique	UX is improved it's seems so compatible to use	-
Unique	Better filtration according to interests	-
Unique	No people just autonomous system	-
Unique	Up to date within minutes of times	-
Unique	Page3 Aim and objectives III News	-

3,160 results	The latest news headlines and articles from a variety of hand-picked sources	Get Free Access
290,000,000 results	Weather	Get Free Access
6,230 results	Current weather conditions, forecasts, maps, news, and traffic	Get Free Access
317,000,000 results	Entertainment	Get Free Access
Unique	Based on the former Bing Enterta	
Unique	Page1 Introduction Problem summary How many page people will choose, it's tough to	
Unique	people and pages are connected to, patience with scrolling to choose pages to follow, and	
1 result	And for which pages they will choose, well, people will pick pages they want	Get Free Access
1 result	very entertaining, interesting, informative, content that reflects their personality, that they like to share, or	Get Free Access
Unique	It is a completely rewritten and redesigned portal, making use of the modern design	
Unique	Much of the existing content on TIMESWEN is eliminate as the website is simplifying	
Unique	Page2 Scope II Trend analysis but it s is better find what s new	
Unique	Content searching of post there, also it shows 1 - 2 minutes ago post	
Unique	TV, movies, music, and celebrity news, as well as theatre show times, tickets, and	

Top plagiarizing domains: [cheesecake.com](#) (1 matches); [Free Download | Mozilla Firefox® Web Browser www.mozilla.org](#) [Download Firefox - the faster, smarter, easier way to browse the web and all of Yahoo 12345Next16,900,000 results](#) (1 matches); [thedailymeal.com](#) (1 matches); [profiles.nlm.nih.gov](#) (1 matches); [cookstr.com](#) (1 matches); [thehealthyfoodie.com](#) (1 matches); [realsimple.com](#) (1 matches);

Create a FREE account to continue.

Page1 Introduction Problem summary I How many page people will choose, it's tough to say? It will depend on the user, how satisfied with their News Feed, how many people and pages are connected to, patience with scrolling to choose pages to follow, and willingness to tinker after they've made an initial selection. All of which is to say that there will probably be a wide range. And for which pages they will choose, well, people will pick pages they want to see posts from. Easy, no? Okay, to be a little less broad, it will be content that people find very entertaining, interesting, informative, content that reflects their personality, that they like to share, or that they think will make them a better person. What is TIMESWEN? TIMESWEN is a web portal and related collection of Internet things & services. It is a completely rewritten and redesigned portal, making use of the modern design language. No employing editors to repurpose existing content from partners at popular organizations. Much of the existing content on TIMESWEN is eliminated as the website is simplifying into a new home page and categories, most of which have corresponding apps. Page2 Scope II Trend analysis but it's better find what's new around the world. So, it should be to Improve trending. Also, rich content after that from another resource & check purity of their contents. UX is improved it seems so compatible to use. Better filtration according to interests. No people just autonomous system. Content searching of post there, also it shows 1 - 2 minutes ago post relevant and it's fast. Up to date within minutes of times. Page3 Aim and objectives III News: The latest news headlines and articles from a variety of hand-picked sources. Weather: Current weather conditions, forecasts, maps, news, and traffic. Entertainment: TV, movies, music, and celebrity news, as well as theatre show times, tickets, and TV listings.

Based on the former Bing Enterta

Introduction

Problem summary I

How many page people will choose, it's tough to say? It will depend on the user, how satisfied with their News Feed, how many people and pages are connected to, patience with scrolling to choose pages to follow, and willingness to tinker after they've made an initial selection. All of which is to say that there will probably be a wide range.

And for which pages they will choose, well, people will pick pages they want to see posts from. Easy, no? Okay, to be a little less broad, it will be content that people find very entertaining, interesting, informative, content that reflects their personality, that they like to share, or that they think will make them a better person.

What is TIMESWEN?

TIMESWEN is a web portal and related collection of Internet things & services. It is a completely rewritten and redesigned portal, making use of the modern design language. No employing editors to repurpose existing content from partners at popular organizations. Much of the existing content on TIMESWEN is eliminate as the website is simplifying into a new home page and categories, most of which have corresponding apps.

Scope II

- Trend analysis but it's better find what's new around the world. So, it should be to Improve trending.
- Also, rich content after that from another resource & check purity of their contents.
- UX is improved it's seems so compatible to use. Better filtration according to interests. No people just autonomous system.
- Content searching of post there, also it shows 1 - 2 minutes ago post relevant and it's fast. Up to date within minutes of times.

Aim and objectives III

- **News:** The latest news headlines and articles from a variety of hand-picked sources.
- **Weather:** Current weather conditions, forecasts, maps, news, and traffic.
- **Entertainment:** TV, movies, music, and celebrity news, as well as theatre show times, tickets, and TV listings. Based on the former Bing Entertainment service. Also includes the games website for online casual games.
- **Sports:** Up-to-the-minute scores, standings, and headlines from leagues worldwide.
- **Money:** Stock market tickers and watch lists, personal finance, real estate, investments, currency converter, and more.
- **Lifestyle:** Headlines, features, and other content related to style, home & garden, family, smart living, relationships, and horoscopes.
- **Health & Fitness:** Tools and information about weight loss, strength, exercise, nutrition, medicine, and more.
- **Food & Drink:** Recipes, cooking tips, news from chefs, cocktails, and shopping lists.
- **Travel:** Destinations, trip ideas, hotel search, flight search, flight status, and arrivals and departures.
- **Autos:** Research and buying advice, auto-related news, information for enthusiasts, and coverage of auto shows worldwide.
- **Video:** Trending and viral videos, comedy and pop culture, and videos.

Problem specifications IV

'The idea is to make the whole news portal system autonomous that fetches every relevant content and article on the Internet.'

Brief literature review and prior art search (PAS) V

The present invention further comprises a system for discovering story trends. The system comprises a plurality of client devices and a plurality of data sources coupled to a network. The system further comprises a web server operable to receive and transmit data to and from the client devices and data sources. In one embodiment, the web server may be further operable for receiving a request for stories from a user and provide a plurality of stories to the user.

Succession planning becomes a focus of their work

A method for identifying story trends includes identifying a set of words in a fixed size data stream based on a sub word cache, and electronically determining at least one-story trend associated with the set of words and electronically generating a story hash associated with the set of words. The method also includes storing the story hash in a story trend cache and updating the story trend cache according to the story hash, and retrieving one or more popular story topics according to the story trend cache. Machine readable media, including program code that causes execution of a method for generating search results also are described.

Technology and tools required VI

Cloud servers

A cloud server is a logical server that is built, hosted and delivered through a cloud computing platform over the Internet. Cloud servers possess and exhibit similar capabilities and functionality to a typical server but are accessed remotely from a cloud service provider.

NGINX

NGINX is a free, open-source, high-performance HTTP server and reverse proxy, as well as an IMAP/POP3 proxy server. NGINX is known for its high performance, stability, rich feature set, simple configuration, and low resource consumption.

NGINX is one of a handful of servers written to address the C10K problem. Unlike traditional servers, NGINX doesn't rely on threads to handle requests. Instead it uses a much more scalable event-driven (asynchronous) architecture. This architecture uses small, but more importantly, predictable amounts of memory under load. Even if you don't expect to handle thousands of simultaneous requests, you can still benefit from NGINX's high-performance and small memory footprint. NGINX scales in all directions: from the smallest VPS all the way up to large clusters of servers.

Bots

Web Bot is an internet bot computer program whose developers claim is able to predict future events by tracking keywords entered on the internet. It was developed in 1997, originally to predict stock market trends. The creator of the Web Bot Project, Clif High, along with his associate George Ure, keep the technology and algorithms largely secret and sell the predictions via the website.

An Internet bot, also known as web robot, WWW robot or simply bot, is a software application that runs automated tasks (scripts) over the Internet. Typically, bots perform tasks that are both simple and structurally repetitive, at a much higher rate than would be possible for a human alone.

Human based pattern detection

Not only are machines rapidly catching up to - and exceeding - humans in terms of raw computing power, they are also starting to do things that we used to consider inherently human. They can feel emotions like regret. They can daydream. So what is - exactly - that humans still do better than machines?

One thing is clear – being able to recognize patterns is what gave humans their evolutionary edge over animals. How we refine, shape and improve our pattern recognition is the key to how much longer we'll have the evolutionary edge over machines.

Apache Lucene Core

Apache Lucene™ is a high-performance, full-featured text search engine library written entirely in Java. It is a technology suitable for nearly any application that requires full-text search, especially cross-platform.

EdgeRank

EdgeRank is the Facebook algorithm that decides which stories appear in each user's newsfeed. The algorithm hides boring stories, so if your story doesn't score well, no one will see it.

TensorFlow

TensorFlow™ is an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

Neo4j

Neo4j equally exploits both data relationships and data elements, empowering the next generation of breakthrough applications.

Natural language processing

Natural language processing (NLP) is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language corpora.

Cytoscape

Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. Cytoscape core distribution provides a basic set of features for data integration, analysis, and visualization.

Design

Analysis, Design Methodology and Implementation Strategy

Functional requirements I

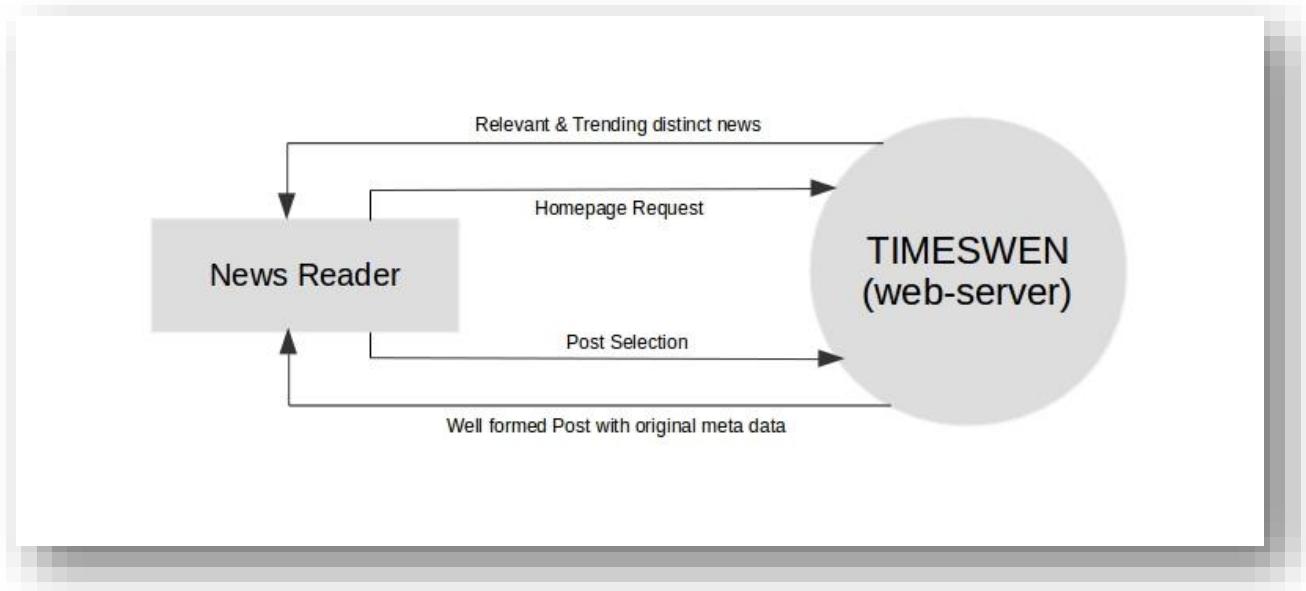
- Web crawling.
- Content parsing.
- Relation mapping.
- Rank and indexing.
- Social media authentication.

Non-functional requirements II

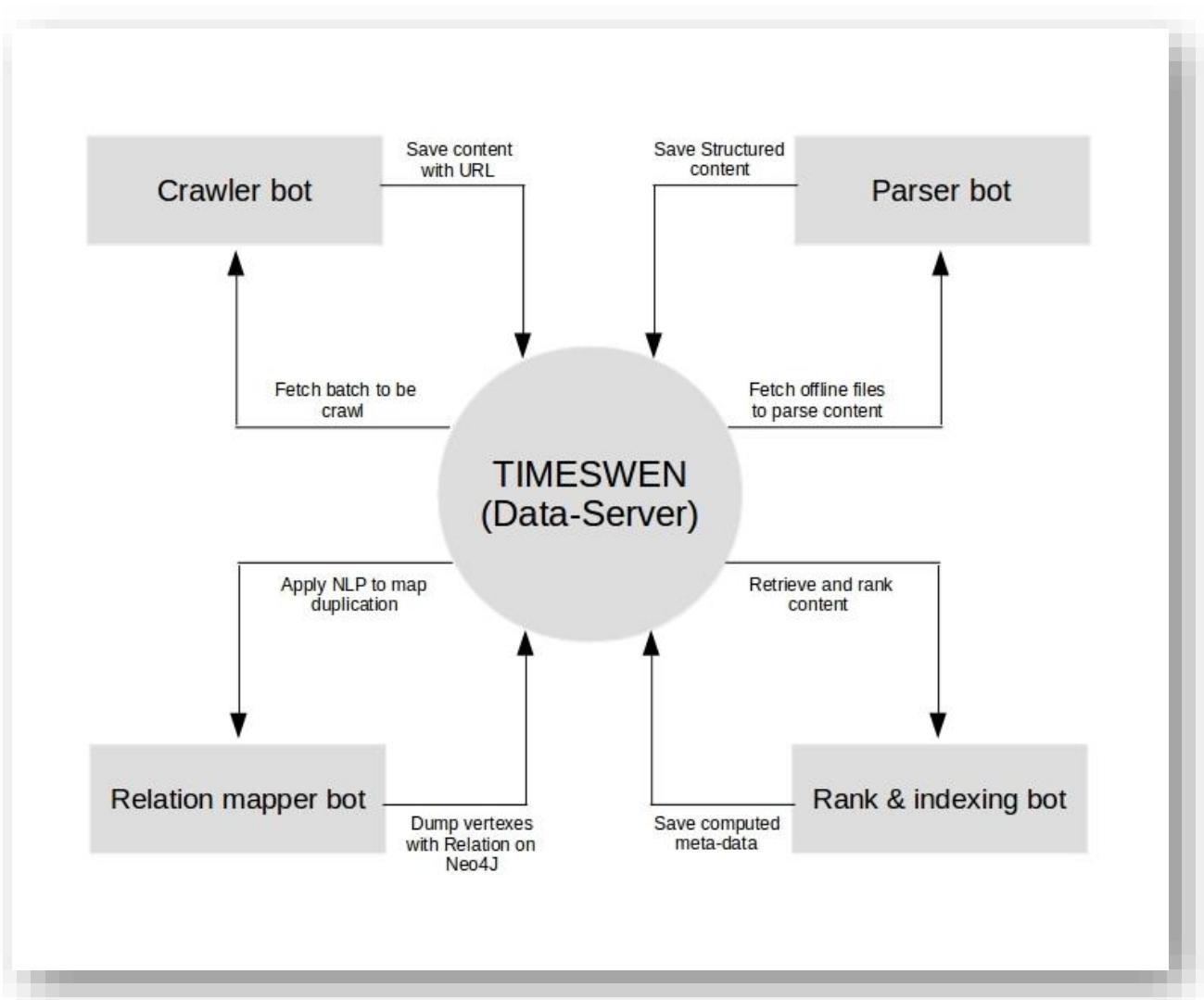
- Load balancer for grand sequence.
- Data integrity.
- Fault toleration.
- Evolution of Eye-Rank.
- High quality content.

Data flow diagram III

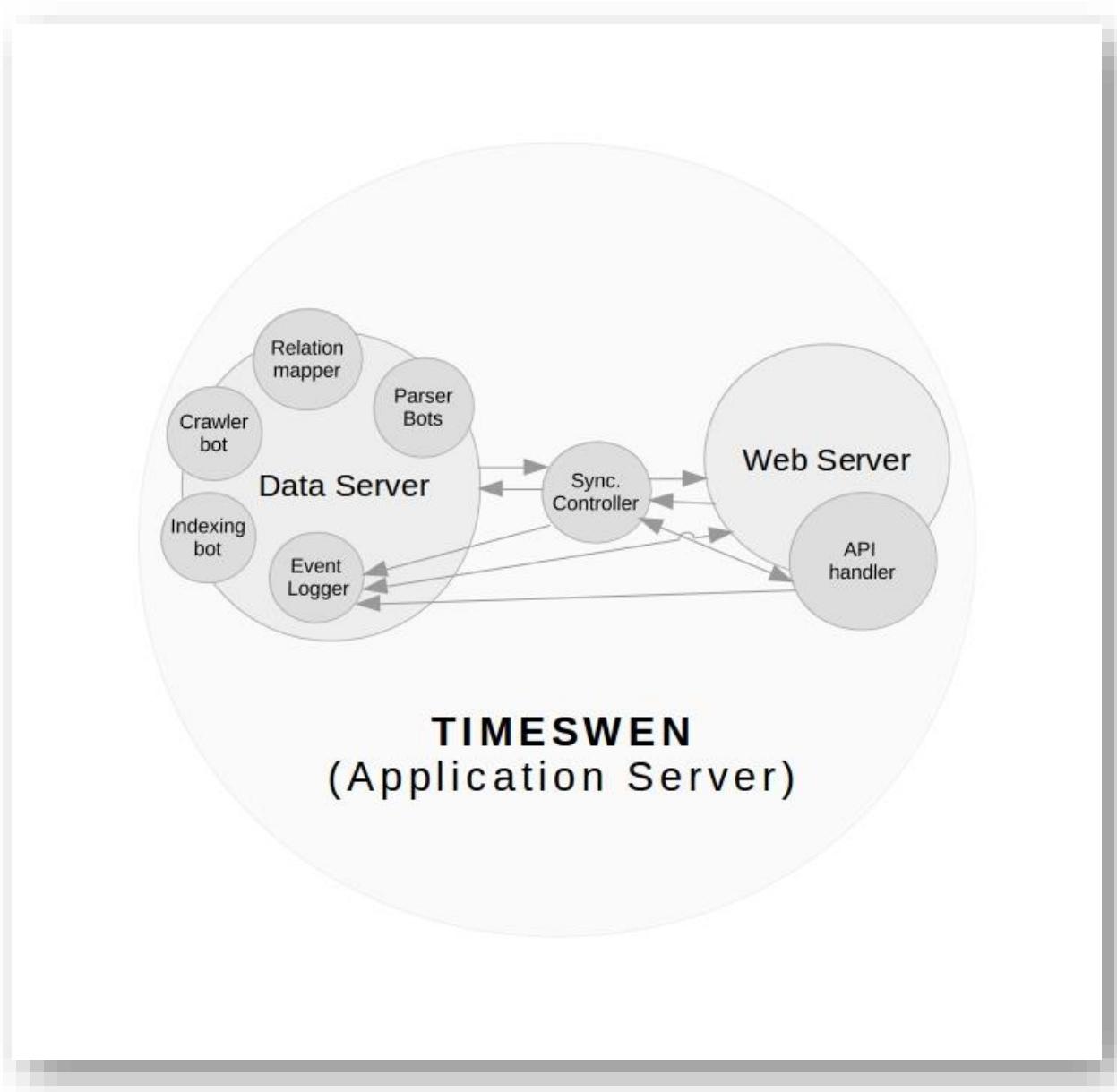
Data flow diagram level 0 – Reader



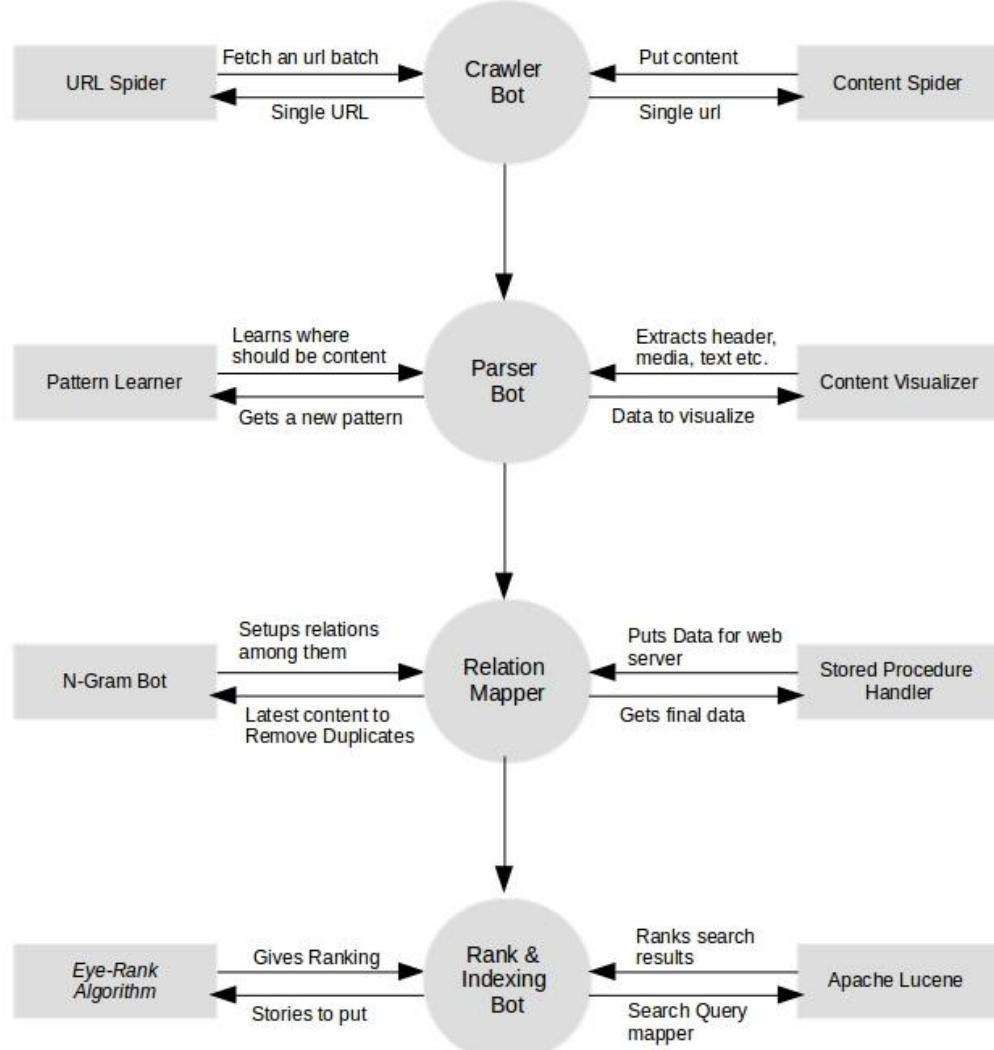
Data flow diagram level 0 - Bots



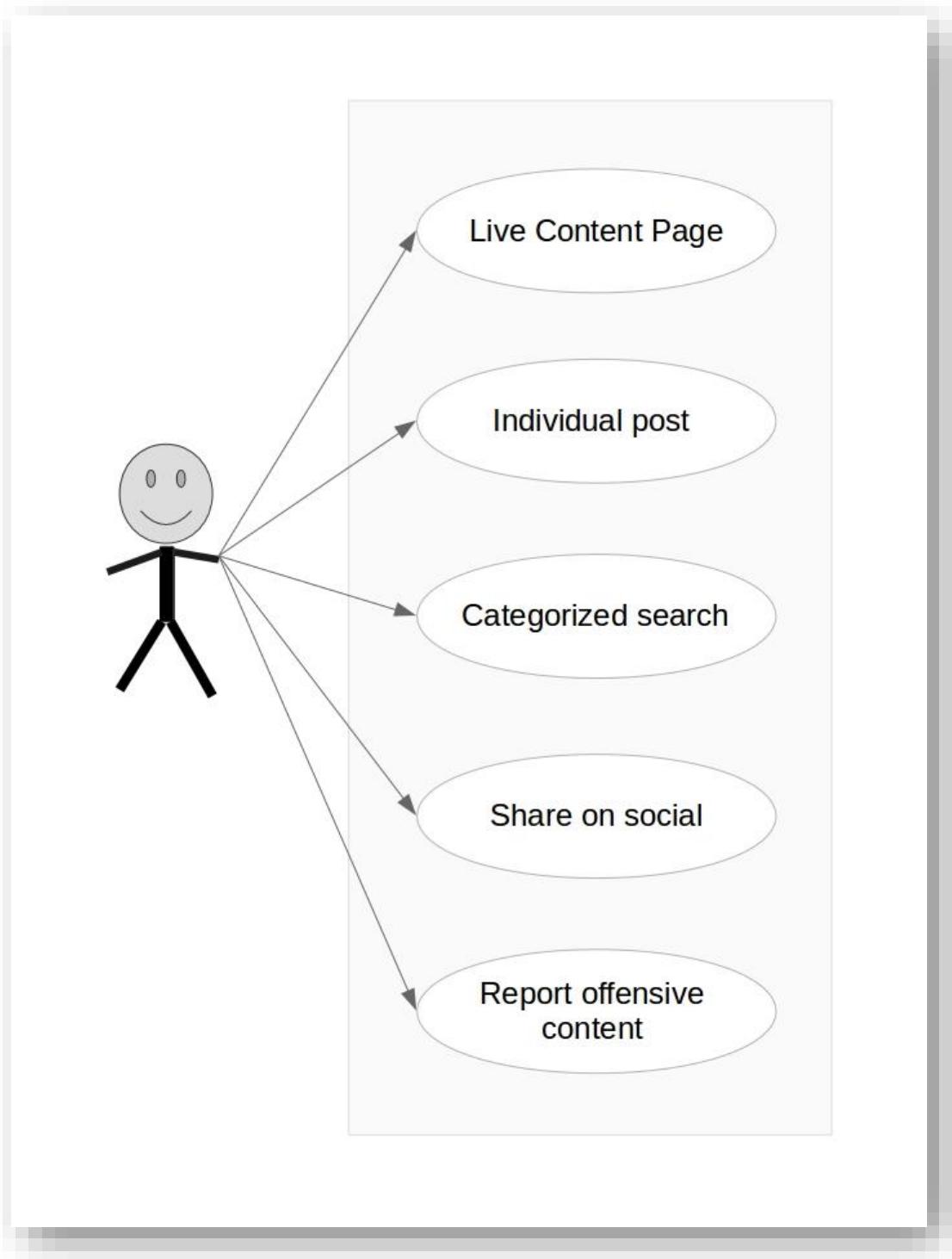
Data flow diagram level 0 - Final overview



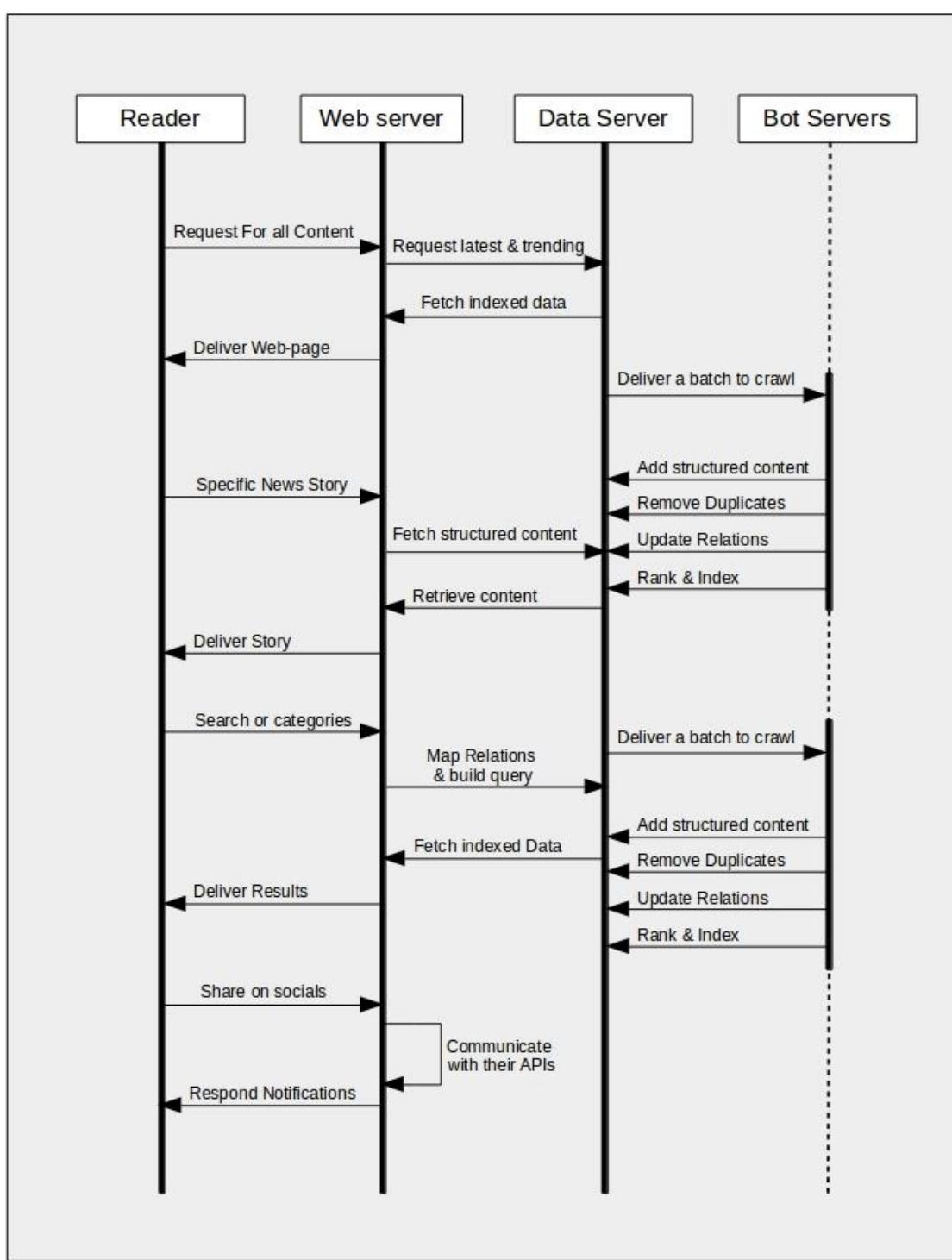
Data flow diagram level 1 - Bots



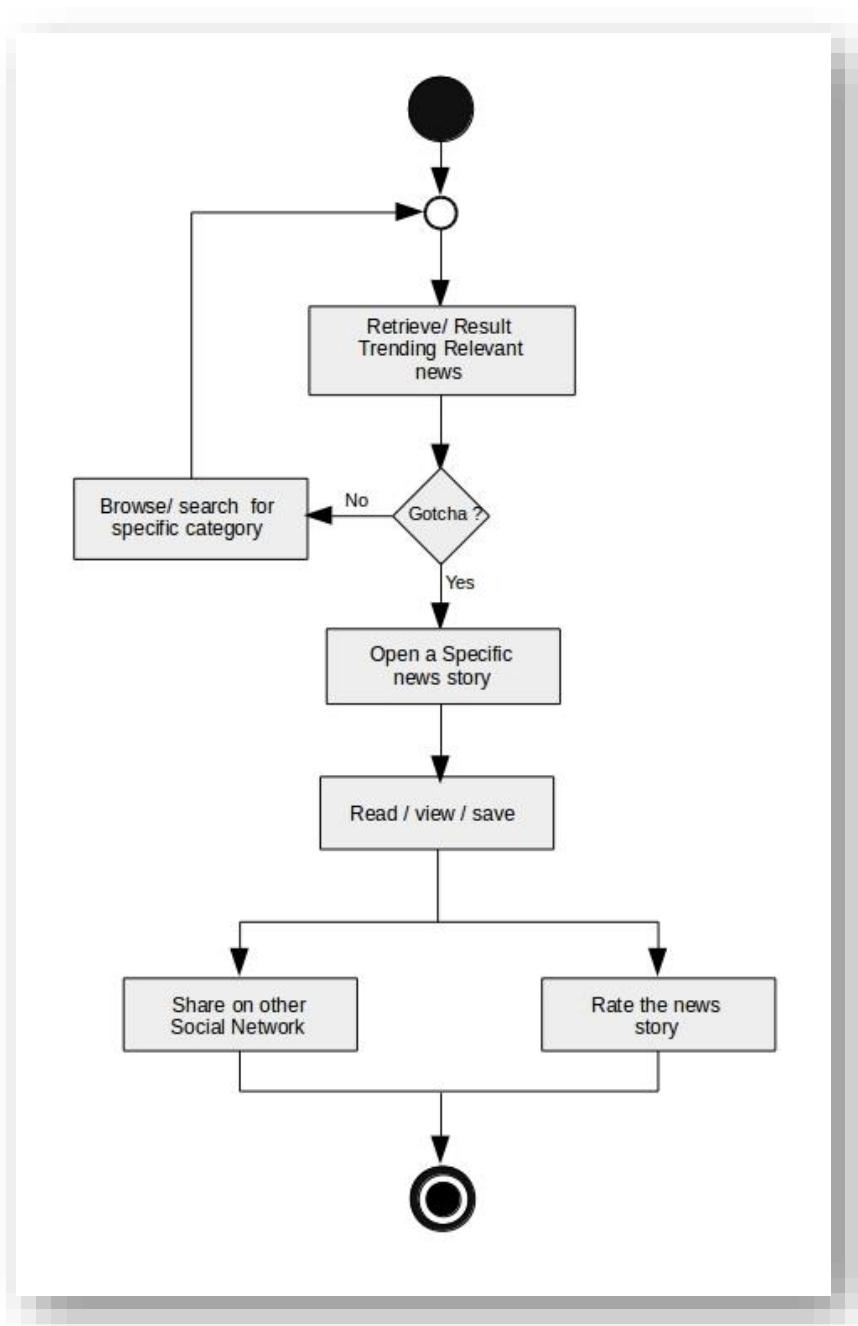
Use case diagram



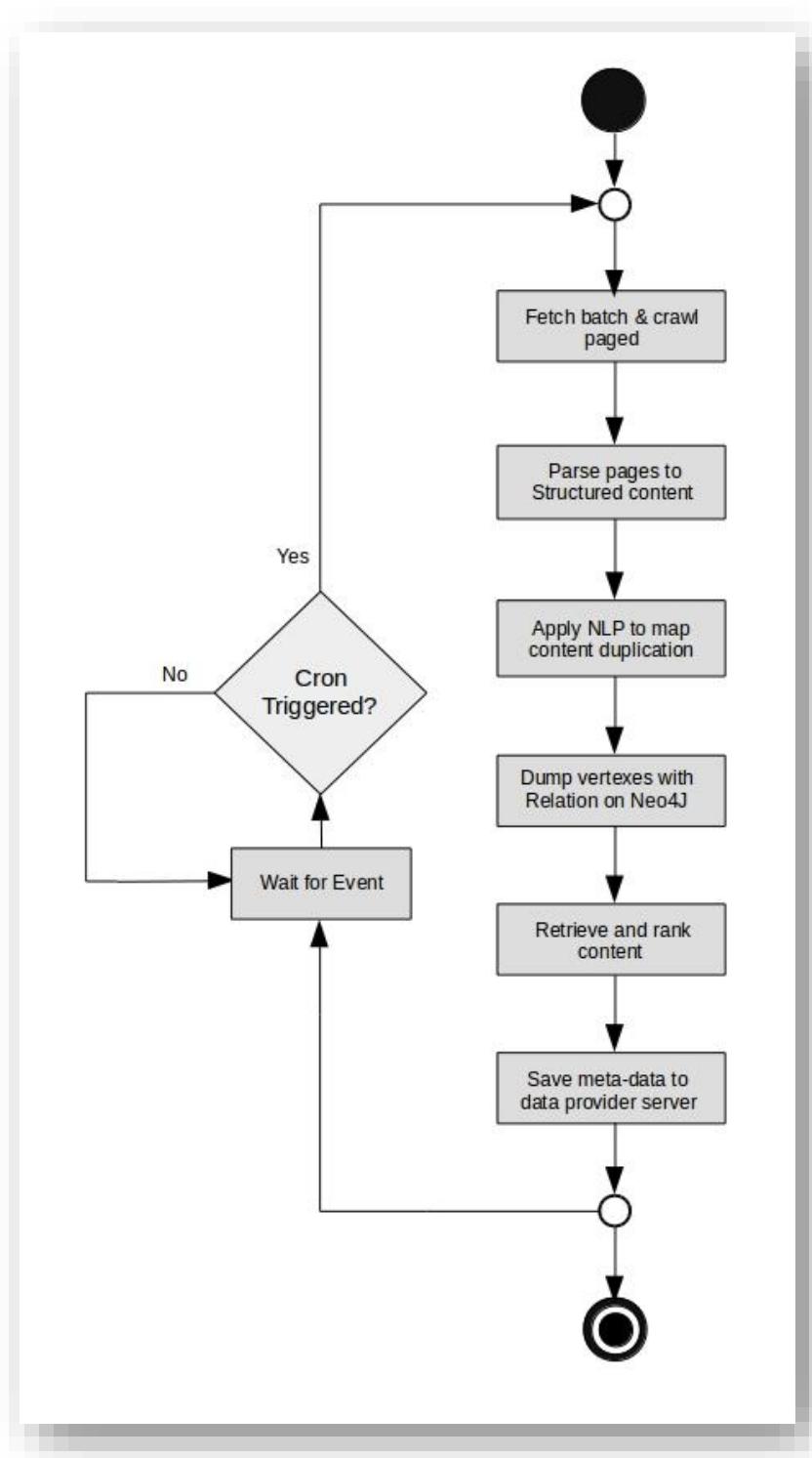
Sequence diagram



Reader's activity diagram



System's activity diagram



Explanation of work V

Observation matrix – AEIOU

AEIOU Framework is a heuristic to help interpret observations gathered by ethnographic practice in industry. Its two primary functions are to code data, and to develop building blocks of models that will ultimately address the objectives and issues of a client. AEIOU stands for 5 elements to be coded: Activity, Environment, Interaction, Object, and User.

- Activities are goal-directed sets of actions – paths towards things people want to accomplish. What are the modes people working in, and the specific activities and processes they go through?
- Environments include the entire arena where activities take place. What is the character and function of the space overall, of each individual's spaces, and of shared spaces?
- Interactions are between a person and someone or something else; they are the building blocks of activities. What is the nature of routine and special interactions between people, between people and objects in their environment, and across distances?
- Objects are building blocks of the environment, key elements sometimes put to complex or unintended uses (thus changing their function, meaning and context). What are the objects, devices people have in their environments, and how do they relate to their activities?
- Users are the people whose behaviors, preferences, and needs are being observed. Who is there? What are their roles and relationships? What are their values and prejudice?

AEIOU Summary:		Group ID: Domain Name:	Date:	Version:
Environment: <ul style="list-style-type: none"> - Festival expressions/ observations - Disaster - Floor plan - Interview features special notes - Scenes 		Interactions: <ul style="list-style-type: none"> - Farmers are writing articles about their interactions with the environment. - Reviewer is interacting with news by typing notes. - User is using sharing features and special notes. 		Objects: <ul style="list-style-type: none"> - Smartphones - Wearables - Tablets/ notebooks - Personal computers
Activities: <ul style="list-style-type: none"> - Reading news stories - Looking for trending topics - Creating photo Summary of news changes languages - Elements, features and special notes - Filtering news by choice or according to categories - Sharing articles and news content on another social media group. - Enabling digital subscriptions - Monitoring stock exchange. 		Users: <ul style="list-style-type: none"> - Researchers - Reporters - Photographers - Businessmen - Journalists - Stock Holders - Celebrities - Storytellers - Sports fans - Politicians - Teenagers - Professors 		

Activities

- Reading news stories.
- Looking for trending topics.
- Different region's users change languages.
- Writing articles on their field of knowledge and researches.
- Enabling digital subscriptions.
- Filtering news by choice or according to categories.
- Sharing articles and news content on another social media group.
- Monitoring stock exchange.

Environment

- Festival.
- Disaster.
- Terrorist activity.
- Election time.
- New movie or album releases.
- Weather forecasts.

Interaction

- Reporter is typing a news.
- Experts are writing articles.
- User is using filtering feature.
- User interacts with helpdesk of the portal.
- Business persons interacts with finance section for relevant news.
- Stock holders interacts with real-time data graphs.
- Stock holders interacts with prediction graphs of stock market.

Objects

- Smartphones.
- Smart watches.
- Personal computers.
- Oculus rift (A VR device).
- Tablets/Notebooks.
- Wearables internet things.

Users

- Reporters.
- Photographers.
- Businessmen.
- Teenagers.
- Stock holders.
- Celebrities.
- Politicians.
- Professors.
- Sportsmen.
- Story readers.
- Researchers.
- Journalist.

Empathy Summary

In this, you have to find out what is user to your problems? What is a stakeholder? What are activities? And what are broad stories of their activities?

User

- Reporters.
- Photographers.
- Businessmen.
- Teenagers.
- Stock holders.
- Celebrities.
- Politicians.
- Professors.
- Sportsmen.
- Story readers.
- Researchers.
- Journalist.

Stakeholders

- Brandings.
- IQ examiners.
- Ad companies.
- Celebrities.
- Publisher.
- Geek persons.
- Organizations.

Activities

- Reading news stories.
- Looking for trending topics.
- Different region's users change languages.
- Writing articles on their field of knowledge and researches.
- Enabling digital subscription.
- Filtering news by choice or according to categories.
- Sharing articles and news content on another social media group.
- Monitoring stock exchange.

Story boarding

Happy

'Since when I started using the service portal, I gain extreme level of general knowledge and became a smart and updated about the happenings around the world.'

'I was new to business world. I used the portal for reading articles about the suggestions and guidelines for new business, as well as the stock market news helped me to purchase new stocks of companies.'

Sad

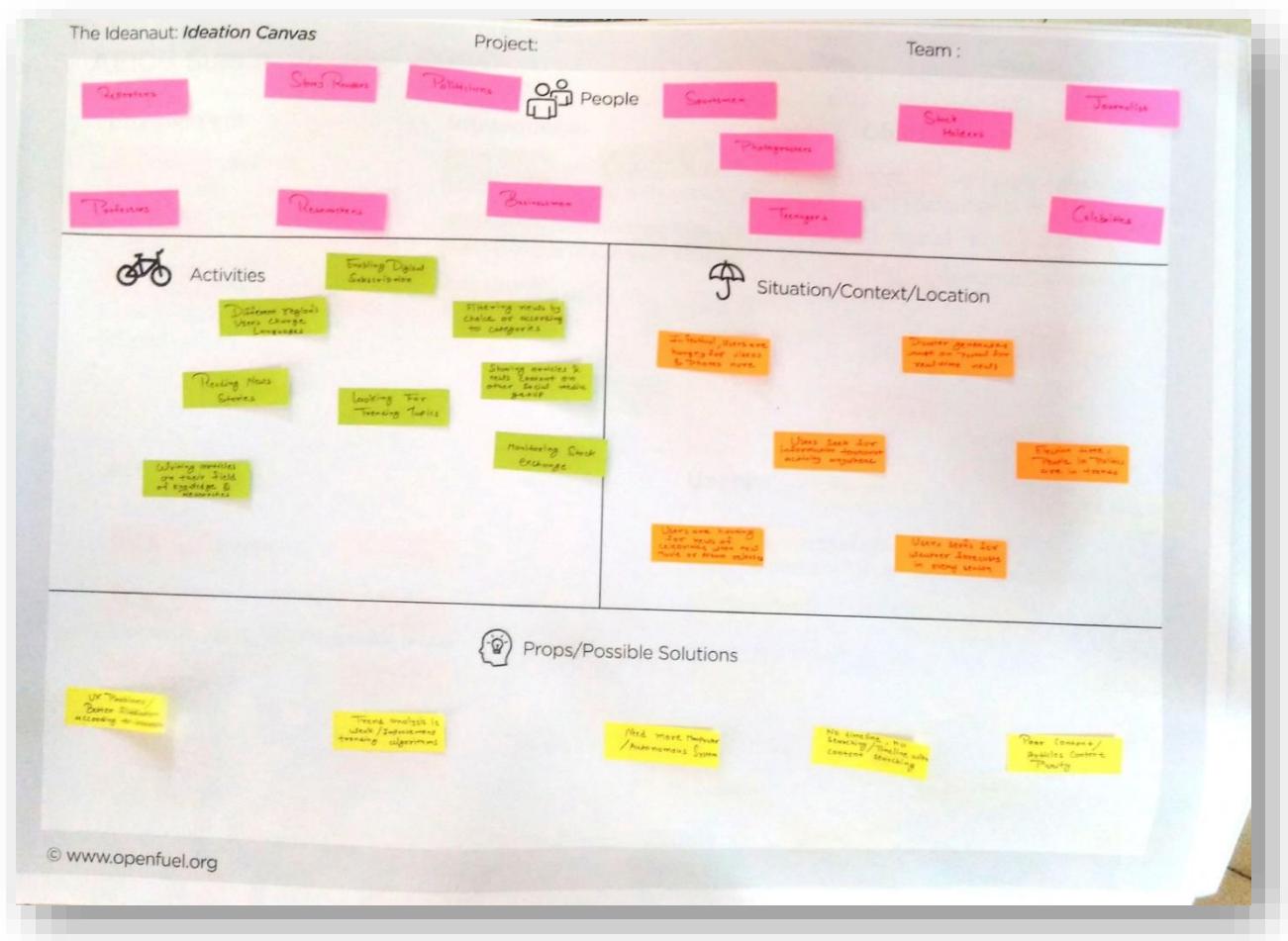
'I read a news article and in excitement I shared in my friend circle. After two days, it was reviled that the news was totally fake. So, I became fool in my friend circle.'

'Rapidly growing internet and digital media became a big threat for the print media as some of newspapers companies have to shut down discard their printing press.'

Ideation canvas

In ideation canvas, you have to carry out which type of activities is related to your project and people? What is situation and location regarding to activities? Then after you find the possible solutions. It is depending or not depend to your activities.

- Who may be the various set of people benefited by your idea?
- Who may be the various stakeholders to your concept?
- Who might use your product?
- Who might invest in your product?



People

- Reporters.
- Photographers.
- Businessmen.
- Teenagers
- Stock holders
- Celebrities
- Politicians
- Professors
- Sportsmen
- Story readers
- Researchers
- Journalist

Activities

- Reading news stories.
- Looking for trending topics.
- Different region's users change languages.
- Writing articles on their field of knowledge and researches.
- Enabling digital subscription.
- Filtering news by choice or according to categories.
- Sharing articles and news content on another social media group.
- Monitoring stock exchange.

Situation/Context/Location

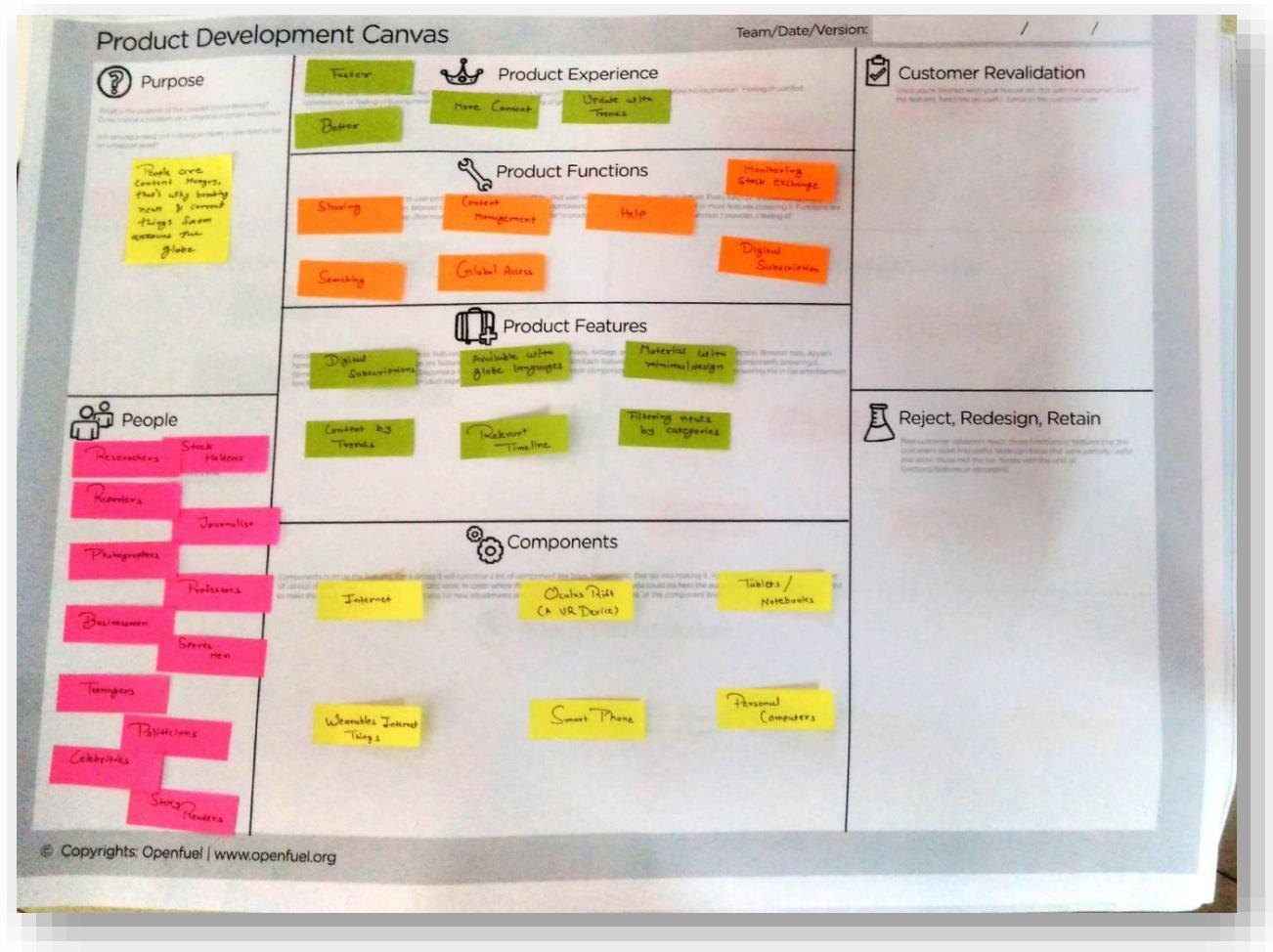
- In festival, users are hungry for videos and photos more.
- Disaster generates rush on portal for real time news.
- Users seek for information terrorist activity anywhere.
- Election time, people in politics are in trends.
- Users are hunting for news of celebrities when new movie or album releases.
- Users serfs for weather forecasts in every season.

Props/Possible Solutions

- Trend analysis is weak/Improvement trending algorithms.
- Poor content/Articles content purity.
- UX problems/Better filtration according to interests.
- Need more manpower/Autonomous system.
- No timeline, no searching/Timeline with content searching.

Product development canvas

The development of the product. From possible solutions, you have idea about what is product? In this canvas, following things is to do.



Purpose

- People are content hungry, that's why breaking news and current things from around the globe.

People

- Reporters.
- Photographers.
- Businessmen.
- Teenagers.

- Stock holders.
- Celebrities.
- Politicians.
- Professors.
- Sportsmen.
- Story readers.
- Researchers.
- Journalist.

Product experience

- Better.
- Faster.
- More content.
- Update with trends.

Product functions

- Sharing.
- Content management.
- Searching.
- Global access.
- Help.
- Digital subscriptions.
- Monitoring stock exchange.

Product features

- Digital subscriptions.
- Available with globe languages.
- Content by trends.
- Relevant timeline.
- Material with minimal design.
- Filtering news by categories.

Components

- Internet.
- Oculus Rift (A VR device).
- Tablets/Notebooks.
- Personal computers.
- Smart phone.
- Wearables Internet things.

Implementation

Performance I

Response time

Mysql vs Neo4j

- 1M users
- Friends of friends for 1K users

Depth	Execution Time – MySQL	Execution Time – Neo4j
2	0. 016	0. 010
3	30. 267	168
4	1, 543. 505	1. 359
5	Not Finished in 1 Hour	2. 132

<http://www.neotechnology.com/how-much-faster-is-a-graph-database-really/>
<http://www.manning.com/partner/>

alberto@graphenedb.com | @albertoperdomo

Data Model	Performance	Scalability	Flexibility	Complexity	Functionality
Key–Value Store	high	high	high	none	variable (none)
Column-Oriented Store	high	high	moderate	low	minimal
Document-Oriented Store	high	variable (high)	high	low	variable (low)
Graph Database	variable	variable	high	high	graph theory
Relational Database	variable	variable	low	moderate	relational algebra

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

TREC Filtering OHSUMED Data Set

63 Topics = Queries ("37 yr old man with sickle cell disease"); Avg. Len: 6.7; OR'ed

196,403 Medical Results (**300MB** Indexable Text)

Judgement Data: (Topic, Result, 2 or 1 or 0 Rating)

Relevancy: **DCG 10**

Platform	Index Peak Memory	Index Time	Index Size	Search Peak Memory	Search Time	Relevancy
Lucene 2.4.1	37 MB	2m15s	91 MB	18 MB	0.02168s (1.366s)	1.0449
zettair 0.9.3	22 MB	0m29.34s	122 MB	9 MB	0.02609s (1.644s)	0.8299
sphinx 0.9.8.1	19 MB	0m42.35s	201 MB	16 MB	0.00803s (0.506s)	0.7690
sqlite 3.6.11	8 MB	1m54.91s	474 MB	7 MB	0.91451s (54.614s)	0.0166
Xapian 1.0.13	48 MB	6m38.17s	339 MB	1 MB	0.02286s (1.440s)	1.0162

Reliability and resource utilization

Since our system will be hosted on a cloud, it will derive the benefits like elastic growth and other new capabilities offered by cloud computing. The following usage scenarios, organized from the simplest to the most complex, illustrate a variety of advantages of virtualization.

- **Hardware independence** – Virtualization minimizes the dependence of an application on its underlying hardware. That means the application may still require the same machine instructions, such as Intel. But its software is decoupled from hardware-based details, such as physical memory and storage, so the application can be easily moved onto new hardware.
- **Server consolidation** – In this case, virtualization increases resource utilization because multiple applications can share hardware resources, including previously underutilized hardware.
- **Multi-tenant** – With this usage scenario, multiple independent instances of an application, such as e-mail or web service, can be consolidated on a single virtualized platform. The instances are then available simultaneously to diverse user communities.
- **Virtual appliance** – In this vision of virtualization, defined by the Distributed Management Task Force (DSP2017), applications are delivered as turnkey software, pre-packaged with operating systems, protocol stacks and supporting software. This approach allows suppliers to thoroughly test the production configuration of all system software, while customers enjoy simpler installation and maintenance.
- **Cloud deployment** – This usage scenario includes rapid elasticity and is the typical endpoint of the evolution of an application to the cloud. It offers the most flexible configuration, which can expand or contract automatically in response to changing workloads.

Design for ergonomics II

For making the Web system more interactive and user friendly, we're focusing on GUI guidelines provided by Google itself called material design which is more about a visual language for our users that synthesizes the classic principles of good design with the innovation and possibility of technology and science. A single underlying system that allows for a unified experience across platforms and device sizes.

- A material metaphor is the unifying theory of a rationalized space and a system of motion. The material is grounded in tactile reality, inspired by the study of paper and ink, yet technologically advanced and open to imagination and magic.
- The foundational elements of print-based design – typography, grids, space, scale, color, and use of imagery – guide visual treatments. These elements do far more than please the eye.
- Each Motion is meaningful and appropriate, serving to focus attention and maintain continuity. Feedback is subtle yet clear. Transitions are efficient yet coherent.

Pattern pages

The screenshot shows the 'Ti Pattern' application interface. At the top, there's a URL bar with 'timeswen.com/home/homepage.php'. Below it, the title 'Ti Pattern' is displayed. A sub-header 'https://www.movzio.com/' is shown above a form. The form has a dropdown menu set to '2' and a text input field labeled 'Type'. Below this is a code snippet: 'Depth 2 : elm[id/class] ~> elm[id/class] => [plaintext/innerhtml/src(href)]'. Underneath, two sections are labeled 'element 1' and 'element 2': 'h3[class=entry-title td-module-title]' and 'a'. To the left of 'element 1' is 'attribute 1' with 'plaintext'. To the right of 'element 2' is a dropdown menu set to 'header/title' with a '+' button. Below this is a section titled 'Ways to crawl { 2 }' containing two items: '* h3[class=entry-title td-module-title] ~> a=>href====>link.href' and '* h3[class=entry-title td-module-title] ~> a=>plaintext====>header/title'. Each item has a blue download icon and a red 'X' icon.

This screenshot shows a more detailed view of the application. It features a large text area containing a complex crawling path or analysis report. The text is a multi-line string of URLs and descriptive text, indicating a sequence of steps or findings from a crawl. At the bottom of this text area is a green 'SUBMIT >' button.

Home pages

TIMESWEN



21 MAY

SCOOPWHOOP

**This Tribute To Dead Police
Mare Shaktiman Is The Most
Ridiculous Thing You Will...**

Just Now 2048 reads



28 APR

SPOTBOYE

**Bipasha kickstarts her
wedding celebrations with a
pooja**

The actress, who will tie the knot with Karan Singh Grover on April 30, had her first prenuptial ceremony

Just Now 524 reads



28 APR

THE VERGE

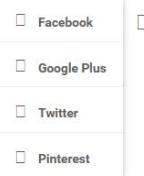
**NASA now uses giant
touchscreens to teach
astronauts how to fly to...**

Just Now 118 reads

Content page

You are done.

Fired. Do not show your face at the laundry again. Stay away from Pinkman. Do not
Are you listening to me?



What do you say?

For now, but he'll come around. In the meantime there's the matter of your brother-in-law. He is a problem you promised to resolve. You have failed. Now is left to me.



To deal with him. If you try to interfere, this becomes a much simpler matter. I will kill your wife, I will kill your son, I will kill your infant daughter.

You are a wealthy man now. One must learn to be rich. To be poor, anyone can manage.
- Gus Fring

photo

fruit

Summary of the results

Problem summary I

- There are lots of those articles are not trustworthy, but still remains in trending, so system finds similarity between articles with advantage of the library collection.
- Creating news graph for identifying relations for existing news articles.

Usefulness

- System train itself so better accuracy in prediction of category and also in trending news articles.
- Evolution of graph of data to knowledge graph.
- Improving content extractor (parser bot) capabilities for any random news.

Benchmarking

- The system is totally dependent on the news sources. If news source is authorized than it has high quality.

Benefits of clustering system

- Reduced single points of failure through Exchange Virtual Server (EVS) failover functionality.
- Ability to perform maintenance and upgrades with limited downtime.
- Ability to easily scale up your cluster to a maximum of seven active EVSs.

References

Microsoft TechNet

[https://technet.microsoft.com/en-us/library/aa996423\(v=exchg.65\).aspx](https://technet.microsoft.com/en-us/library/aa996423(v=exchg.65).aspx)

NGINX

<https://www.nginx.com>

Neo4j

<https://neo4j.com>

Apache Lucene Core

<https://lucene.apache.org/core/>

TensorFlow

<https://www.tensorflow.org>

Cytoscape

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

EdgeRank

<http://edgerank.net/>

PRESENTING

EdgeRank

A GUIDE TO FACEBOOK'S NEWSFEED ALGORITHM

fig 2.1
Source: EdgeRank.net

$$\sum_{edges\ e} u_e w_e d_e$$

U_e ~ affinity score between viewing user and edge creator
 W_e ~ weight for this edge type (create, connect, like, tag, ect.)
 d_e ~ time decay factor based on how long ago the edge was created

What is EdgeRank?

EdgeRank is the Facebook algorithm that decides which stories appear in each user's newsfeed. The algorithm hides boring stories, so if your story doesn't score well, no one will see it.

The first thing someone sees when they log into Facebook is the newsfeed. This is a summary of what's been happening recently among their friends on Facebook.

Every action their friends take is a potential newsfeed story. Facebook calls these actions "Edges." That means whenever a friend posts a status update, comments on another status update, tags a photo, joins a fan page, or RSVP's to an event it generates an "Edge," and a story about that Edge might show up in the user's personal newsfeed.

It'd be completely overwhelming if the newsfeed showed all of the possible stories from your friends. So Facebook created an algorithm to predict how interesting each story will be to each user. Facebook calls this algorithm "EdgeRank" because it ranks the edges. Then they filter each user's newsfeed to only show the top-ranked stories for that particular user.

Why should I care?

Because most of your Facebook fans never see your status updates.

Facebook looks at all possible stories and says "Which story has the highest EdgeRank score? Let's show it at the top of the user's newsfeed. Which one has the next highest score? Let's show it next." If EdgeRank predicts a particular user will find your status update boring, then your status update will never even be shown to that particular user.

Caveat: There actually appears to be two algorithms, although this has not been conclusively proven. The EdgeRank algorithm ranks stories, and a second algorithm sorts the newsfeed. This newsfeed algorithm includes a randomization element and a [keyword aggregator](#). Zuckerberg mentioned in an interview with TechCrunch that Facebook users found it every how well Facebook knew what they were interested in, so they started randomizing the newsfeed slightly.

The numbers on this are frightening. In 2007, a Facebook engineer said in an interview that only about [0.2% of eligible stories make it into a user's newsfeed](#). That means that your status update is competing with 499 other stories for a single slot in a user's newsfeed.

How does EdgeRank work?

EdgeRank is like a credit rating: it's invisible, it's important, it's unique to each user, and no one other than Facebook knows exactly how it works.

At Facebook's [2010 F8 conference](#), they revealed the three ingredients of the algorithm:

News feed FYI: More articles you want to spend time viewing

As we work to improve News Feed, we make updates to help make sure you see the most relevant stories at the top. As part of this work, we ask thousands of people to rate their experience every day and tell us how we can improve what they see when they check Facebook — we call this our Feed Quality Program. From these conversations, we have learned that the actions people take on Facebook—liking, clicking, commenting or sharing a post—don't always tell us the whole story of what is most meaningful to them.

For example, we've found that there are stories people don't like or comment on that they still want to see, such as articles about a serious current event, or sad news from a friend. Based on this finding, we previously updated News Feed's ranking to factor in how much time you spend reading a post within News Feed, regardless of whether you opened the article. We also previously updated News Feed's ranking to take into account times when someone clicked on an article and came straight back to News Feed as we learned that this often happened when the article someone clicked on wasn't what they had expected from the post or the headline.

Time spent viewing

Building on this work, we're learning that the time people choose to spend reading or watching content they clicked on from News Feed is an important signal that the story was interesting to them. We are adding another factor to News Feed ranking so that we will now predict how long you spend looking at an article in the Facebook mobile browser or an Instant Article after you have clicked through from News Feed. This update to ranking will take into account how likely you are to click on an article and then spend time reading it. We will not be counting loading time towards this — we will be taking into account time spent reading and watching once the content has fully loaded. We will also be looking at the time spent within a threshold so as not to accidentally treat longer articles preferentially.

With this change, we can better understand which articles might be interesting to you based on how long you and others read them, so you'll be more likely to see stories you're interested in reading. This change only factors in the time people spend reading an article regardless of whether that time is spent reading an Instant Article or an article in the mobile web browser.

Diversity of page posts

We've also heard from people that they enjoy reading articles from a wide range of publishers, and it can be repetitive if too many articles from the same source are back to back in their News Feed. We'll also be making an update to reduce how often people see several posts in a row from the same source in their News Feed.

Will this impact my page?

We've started rolling this out and will continue over the coming weeks. We anticipate that most Pages won't see any significant changes. Some Pages might see a small increase in referral traffic, and some Pages might see minor decreases. As always, Pages should refer to our publishing best practices.

How does news affect stock prices?

<http://www.sharemarketschool.com/how-does-news-affect-stock-prices/>

News is something that affects stock prices. Whether you're a long-term investor or a short-term investor, it's important to review the news headlines periodically. There may be positive news, negative news or news to which market may not react at all. One has to be smart enough to decode the news and quickly grasp whether it will affect his stocks in anyway and if yes, the degree to which the news can have an impact.

Positive and negative news

News which is considered as positive tends to have a positive effect on stock markets and one can see share prices rising soon after the news come out in the open. Positive news — such as a joint venture agreement, securing of new orders, healthy sales numbers, discovery of huge oil reserves in a country, excellent financial results of accompany etc. should send a stock up. Stock prices react slowly but steadily to positive news.

- A curious fact that I have noted in some cases is that good news doesn't always translate to a jump in stock price; in fact, often the good news will produce a slight drop in a stock price. Why? Because unofficial news, also known as "rumors", can have as much impact on stock prices as official news announcements. The stock market often anticipates these news stories and "prices in" its expectations accordingly. When those expectations are confirmed with actual investment news, the price may temporarily drop. Of course, the reverse applies, too: if rumors swirling around a stock aren't proven true, investors may respond in surprising ways. If the surprise is a good one, stock prices can be driven upward as a result. That's why it's key to watch the investment news online and see how headlines influence stock quotes.
- Another point observed is that good news at home and bad news abroad can adversely push stock prices down. The international market is intertwined within the home market. Sometimes, all it takes is a bit of bad news from overseas to have a down market day.

Negative news has more far reaching effect on stock prices and investor sentiment than positive news. Stock prices react very heavily to negative news that it may seriously stop average people from wanting to buy stocks.

The sentiment of the market is also an important factor. In a largely negative atmosphere, the slightest bit of worrisome news is enough to send a stock tumbling.

'Good' bad news

There is some news which might seem negative at first but it isn't actually negative. For example, firing of CEO or top officials. This may sound very negative at first, but it does show that the company's board of directors was bold enough to take drastic actions to help the company in the long run. Another example is lay-offing in a company. This is usually good for the company and its stock price because expenses will be reduced significantly and quickly. This should help increase earnings right away. It is not always a major warning sign; it could just be a reaction to a slower economy. It is one of the quickest ways a company can cut expenses if sales have not been meeting expectations.

Stock blogs

Apart from news that comes in Medias like television channels and newspapers which brings out the actual news, Blog is another category that influences stock market investors. The difference between a blog and other Medias is that a Blog is usually maintained by an individual with regular entries of commentary. Most of them contain opinions on a particular event along with the actual news. So just in case you are confused about what's going to happen in stock markets after RBI hikes interest rate or what's in store for apple after Steve jobs era, these are sources to know what the people think about it following some good blogs on stock markets would help you to understand economic events better.

Conclusion

The news has a direct impact on the market. It can change a bad day into a good one or a good day into a bad one. The relationship between the news and the market can be highly unpredictable by the best analysts. The next headline can turn out to be a boon or a bust.

- Good news will have a positive impact on stock prices
- Stock prices reacts to negative news quickly than it would react to a positive news.
- The good news locally can be overshadowed by the negative news across the globe.
- In a negative atmosphere, the slightest bit of worrisome news is enough to send a stock tumbling. The opposite is also true. To an extent, news effects largely depend on the reigning sentiment rather than the actual significance of the news.
- Just because the news is bad doesn't mean the stock market will have a bad day.
- News about the following affects stock markets: Crude Oil prices, IIP, Inflation, Unemployment, government policies, political unrest, draught or monsoon, company results, Global cues, FII activities, mergers and acquisitions , insider trading, bonus dividends and stock buy backs, stock splits, rights issue, inclusion or exclusion from indexes, change or death of top officials, loss of customers or break through deals, changes in demand and supply, fluctuations in prices of raw materials, war, terrorist attacks, joint ventures, rumors , new interventions etc.

Referenced content



<http://www.msn.com/en-in/entertainment/hollywood/deepika-padukone-priyanka-chopra-in-the-race-for-the-next-bond-girl/ar-BBsRg1Z>

The screenshot shows a news article from the MSN Entertainment section. The headline reads "Deepika Padukone, Priyanka Chopra in the race for the next Bond girl?". The article discusses the possibility of Indian actresses Deepika Padukone and Priyanka Chopra replacing Daniel Craig as James Bond. It includes a photo of the two actresses and a sidebar with a Microsoft Azure advertisement and a "YOU MAY LIKE" section.

msn entertainment

Recent Searches: Deepika Padukone, Priyanka Chopra in the r...

Sign in | Manage History

Home | Bollywood | TV | South Cinema | Hollywood | Reviews | Photos | Video | 'Traffic' | 'Game Of Thrones' | Awards | Quiz | Specials

Bhaaiyon, Behroni! Bing Karo, Fir Vote kar!

#AssemblyElections

Deepika Padukone, Priyanka Chopra in the race for the next Bond girl?

TODAY

India Today

1 day ago

SHARE | TWEET | SHARE | EMAIL

Deepika Padukone and Priyanka Chopra are both creating waves overseas. While Deepika has got the temperature soaring with 'XXX The Return Of Xander Cage', Priyanka has raised the bar by winning the villain's part in 'Baywatch'. And even as we in India celebrate the Desi Girls' achievements abroad, the two are giving us new reasons to cheer for them with every passing day or so it seems.

© Provided by India Today

If the latest reports from that part of the world are to be believed, Deepika and Priyanka are both supposedly being considered for the role of the next Bond girl. The last in the James Bond franchise, 'Spectre', hit the screens in 2015.

While the brains behind Bond are still trying to zero in on an apt replacement for Daniel Craig, it seems like Deepika and Priyanka are both trying hard to land the dream role of the next Bond girl.

Microsoft Azure

Move apps to the cloud without rewriting them. Simple. Flexible. Connected.

Free trial

YOU MAY LIKE

Let Your Salary Work as

Original content



<http://indiatoday.intoday.in/story/deepika-padukone-priyanka-chopra-in-the-race-for-the-next-bond-girl/1/664367.html>

A screenshot of a news article from IndiaToday.in. The header includes a menu icon, the website logo, a "NEWS" tab (with "LIVE TV" and "MAGAZINE" options), and a search bar. The main headline is "Deepika Padukone, Priyanka Chopra in the race for the next Bond girl?". Below the headline is a subtext: "Deepika Padukone and Priyanka Chopra, reports suggest, are vying for the role of the next Bond girl." A photo of the two actresses is displayed. To the left of the photo are social sharing icons for LinkedIn, Facebook, Twitter, and Google+, with a total of "4.39k SHARES". To the right of the photo is an advertisement for "Aarogyam Presents Wellness Packages" offered by Thyrocare, starting from Rs 600*. Below the advertisement is a "Related Stories" section featuring a link to "Dwayne Johnson's 'sistah' love for Priyanka Chopra". At the bottom of the page is a "More From Movies" section.

Patents

13/856398

<http://www.freepatentonline.com/9235635.html>

US09343511

<https://patents.google.com/patent/US6547829B1/en?q=Content&q=Duplication&q=Detection>

14/147789

<http://www.freepatentonline.com/9633119.html>

15326045

<https://patentscope.wipo.int/search/en/detail.jsf?docId=US200949331&recNum=5&tab=Drawings&maxRec=488&office=&prevFilter=&sortOption=Pub+Date+Desc&queryString=F%3A%28web+crawling%29>

US10611269

<https://patents.google.com/patent/US7568148B1/en?q=news&q=duplicate>

Appendix

Periodic progress reports (PPR)

[Print](#) [Back](#)

College : L. D. COLLEGE OF ENGINEERING, AHMEDABAD
StudentName : Akhawat Kalpit Ajeetkumar
EnrollmentNo : 150283116001
MobileNo : 9727331128
Email : kalpitakhawat@gmail.com
Department : Information Technology
Discipline : BE
Semester : Semester 7

PPR Details

Periodic Progress Report : First PPR

Project : Pure Trending Content (Timeswen)

Status : Reviewed

1. What Progress you have made in the Project ?

We have done some analysis process on current system and some on UI and start to learn whatever technology is going to be used.

2. What challenge you have faced ?

How To Get The Data from all Sources?

3. What support you need ?

We required high computation power for parsing and removing duplication from crawled data.

4. Which literature you have referred ?

Some Data Analytics And Natural Language Processing research papers, blogs and journals

Comments

Comment by Internal Guide :

ok

Comment by External Guide :

None

Comment by HOD :

None

Comment by Principal :

None

Comment by University Admin :

None

[Print](#) [Back](#)

College : L. D. COLLEGE OF ENGINEERING, AHMEDABAD

StudentName : Bhatt Jigar Maneshbhai

EnrollmentNo : 150283116002

Department : Information Technology

MobileNo : 9638967123

Discipline : BE

Email : bhatt.jigar.214@ldce.ac.in

Semester : Semester 7

PPR Details

Periodic Progess Report : Second PPR

Project : Pure Trending Content (Timeswen)

Status : Reviewed

1. What Progress you have made in the Project ?

We started researching about parallel processing architectures available for big data analysis. This research brought us on the cloud finally but it was very expensive so we had research on cloud implementation with local machines.

2. What challenge you have faced ?

How to implement cloud like architecture with local machines for achieving high computation power.

3. What support you need ?

Parallel Computation Power & a cloud setup. Hypervisor expert for cloud & cluster guidance.

4. Which literature you have referred ?

<http://www.gtia.co.in/papers/e9bd6c2f-e271-44e6-8ef9-2fc36d79bbc31112012.pdf>

<http://www.communitysense.nl/papers/cacm06b.pdf>

<https://softwareengineering.stackexchange.com/questions/257661/patterns-for-creating-adaptive-web-crawler-throttling> https://www.tensorflow.org/api_docs/ <http://curator.apache.org/>

Comments

Comment by Internal Guide :

None

Comment by External Guide :

None

Comment by HOD :

None

Comment by Principal :

None

Comment by University Admin :

None

[Print](#) [Back](#)

College : L. D. COLLEGE OF ENGINEERING, AHMEDABAD

StudentName : Fotariya Jimish Maheshbhai

EnrollmentNo : 150283116008

Department : Information Technology

MobileNo : 9712252863

Discipline : BE

Email : fotariyajimish@gmail.com

Semester : Semester 7

PPR Details

Periodic Progess Report : Third PPR

Project : Pure Trending Content (Timeswen)

Status : Reviewed

1. What Progress you have made in the Project ?

We are learning about natural language processing & libraries like Natural language toolkit for finding how the machine can extract human language semantics. Were also checking the existing web crawlers & parser structure.

2. What challenge you have faced ?

How to Process data which have the same meaning but in different word structure.

3. What support you need ?

Expert for natural language processing.

4. Which literature you have referred ?

https://www.tensorflow.org/api_docs/

www.ijcscn.com/Documents/Volumes/vol4issue6/ijcscn2014040606.pdf

<https://www.w3.org/standards/semanticweb/> semanticweb.org/ <https://userpages.uni-koblenz.de/~staab/Research/Publications/sub-flairs2002.pdf> <https://docs.microsoft.com/en-us/azure/>

Comments

Comment by Internal Guide :

None

Comment by External Guide :

None

Comment by HOD :

None

Comment by Principal :

None

Comment by University Admin :

None

[Print](#)[Back](#)

College : L. D. COLLEGE OF ENGINEERING, AHMEDABAD

StudentName : Solanki Hardik Dineshbhai

EnrollmentNo : 150283116026

Department : Information Technology

MobileNo : 8000334996

Discipline : BE

Email : hardik_solanki@live.com

Semester : Semester 7

PPR Details

Periodic Progress Report : Forth PPR

Project : Pure Trending Content (Timeswen)

Status : Reviewed

1. What Progress you have made in the Project ?

We've started learning Algorithm Optimization techniques since Ranking of content is going to be done in a frequent way because of an uncertainty of timing in the world of trending news.

2. What challenge you have faced ?

We wanted to speed up our ranking algorithm but there is an issue of algorithm inconsistency. Good-trading content can lose rank because of speed up modification logics in the existing algorithm. So the final decision was taken was to design a new ranking algo. from scratch.

3. What support you need ?

Other experts asked before, plus a GraphQL & neo4j Expert.

4. Which literature you have referred ?

https://www.tensorflow.org/api_docs/ <https://codesachin.wordpress.com/2015/11/14/k-means-clustering-with-tensorflow/> <https://www.tensorflow.uk/tag/hierarchical-clustering/>

<http://mahout.apache.org/docs/0.13.1-SNAPSHOT/algorithms/linear-algebra/d-sPCA.html>

<http://ieeexplore.ieee.org/abstract/document/58871/?reload=true> www.nltk.org/

Comments

Comment by Internal Guide :

None

Comment by External Guide :

None

Comment by HOD :

None

Comment by Principal :

None

Comment by University Admin :

None

Patent search and analysis reports (PSAR)



**GUJARAT TECHNOLOGICAL UNIVERSITY
(GTU)
INNOVATION COUNCIL (GIC)
Patent Search & Analysis Report
(PSAR)**



Date of Submission : 31/08/2017

Dear Akhawat Kalpit Ajeetkumar,

Studied Patent Number for generation of PSAR : 17BE7_150283116001_1

PART 1: PATENT SEARCH DATABASE USED

1. Patent Search Database used : Indian Patent Office database
Web link of database : <http://ipindiaservices.gov.in/publicsearch/>
2. Keywords Used for Search : Trending ,Content,news
3. Search String Used : pure trending content news algorithm
4. Number of Results/Hits getting : 0

PART 2: BASIC DATA OF PATENTED INVENTION /BIBLIOGRAPHIC DATA

5. Category/ Field of Invention :
6. Invention is Related to/Class of Invention : Get Trending Data
6 (a) : IPC class of the studied patent : G06F17/30; G06F7/00; G06F17/27
7. Title of Invention : System and method for discovering story trends in real time from user generated content
8. Patent No. :
9. Application Number : 13/856398
9 (a) : Web link of the studied patent : <http://www.freepatentsonline.com/9235635.html>
10. Date of Filing/Application (DD/MM/YYYY) : 03/04/2013
11. Priority Date (DD/MM/YYYY) :
12. Publication/Journal Number :
13. Publication Date (DD/MM/YYYY) :
14. First Filled Country : Albania

15. Also Published as

Sr.No	Country Where Filed	Application No./Patent No.
1		

16. Inventor/s Details.

Sr.No	Name of Inventor	Address/City/Country of Inventor
1	Ittiachen	Bangalore, IN

17. Applicant/Assignee Details.

Sr.No	Name of Applicant/Assignee	Address/City/Country of Applicant
1	Yahoo Inc	Sunnyvale, CA, US

18. Applicant for Patent is : Individual

PART 3: TECHNICAL PART OF PATENTED INVENTION**19. Limitation of Prior Technology / Art**

Currently, the trends in these user generated streams are surfaced as "ebay", "announce deal", "sell skype", because users write about the same topics differently. The current state of the art fails to cohesively analyze user-generated streams to account for the variance in terminology used across a diverse data set

20. Specific Problem Solved / Objective of Invention

The present invention provides a solution allowing a system to intelligently parse and identify key trending topics and store topics or stories for subsequent analysis and retrieval.

21. Brief about Invention

A method for identifying story trends includes identifying a set of words in a fixed size data stream based on a subword cache, and electronically determining at least one story trend associated with the set of words and electronically generating a story hash associated with the set of words. The method also includes storing the story hash in a story trend cache and updating the story trend cache according to the story hash, and retrieving one or more popular story topics according to the story trend cache. Machine readable media including program code that causes execution of a method for generating search results also are described.

22. Key learning Points

Trending Analysis

23. Summary of Invention

The present invention further comprises a system for discovering story trends. The system comprises a plurality of client devices and a plurality of data sources coupled to a network. The system further comprises a web server operable to receive and transmit data to and from the client devices and data sources. In one embodiment the web server may be further operable to receiving a request for stories from a user and provide a plurality of stories to the user.

24. Number of Claims : 13

25. Patent Status : Published Application

26. How much this invention is related with your IDP/UDP?

Not related to IDP/UDP, It's related to my area of interest

27. Do you have any idea to do anything around the said invention to improve it? (Give short note in not more than 500 words)

No. I don't have any idea about improvement on this. We have ideas of integrating this with other algorithms.



**GUJARAT TECHNOLOGICAL UNIVERSITY
(GTU)
INNOVATION COUNCIL (GIC)
Patent Search & Analysis Report
(PSAR)**



Date of Submission : 11/09/2017

Dear Solanki Hardik Dineshbhai,

Studied Patent Number for generation of PSAR : 17BE7_150283116026_2

PART 1: PATENT SEARCH DATABASE USED

1. Patent Search Database used : Google Patents
Web link of database : <https://patents.google.com/>
2. Keywords Used for Search : Content Duplication detection,Content Duplication detection,Content Duplication detection
3. Search String Used : Content Duplication detection
4. Number of Results/Hits getting : 7373

PART 2: BASIC DATA OF PATENTED INVENTION /BIBLIOGRAPHIC DATA

5. Category/ Field of Invention :
6. Invention is Related to/Class of Invention : 1/1
6 (a) : IPC class of the studied patent : NA
7. Title of Invention : Method and system for detecting duplicate documents in web crawls
8. Patent No. :
9. Application Number : US09343511
9 (a) : Web link of the studied patent : <https://patents.google.com/patent/US6547829B1/en?q=Content&q=Duplication&q=Detection>
10. Date of Filing/Application (DD/MM/YYYY) : 06/30/1999
11. Priority Date (DD/MM/YYYY) :
12. Publication/Journal Number :
13. Publication Date (DD/MM/YYYY) :
14. First Filled Country : Albania

15. Also Published as

Sr.No	Country Where Filed	Application No./Patent No.
1		

16. Inventor/s Details.

Sr.No	Name of Inventor	Address/City/Country of Inventor
1	Dmitriy Meyerzon	NA
2	Srikanth Shoroff F	NA
3	Soner Terek	NA
4	Scott Norin	NA

17. Applicant/Assignee Details.

Sr.No	Name of Applicant/Assignee	Address/City/Country of Applicant
1	Microsoft Technology Licensing LLC	1288 Pear Ave, CA, US

18. Applicant for Patent is : Group

PART 3: TECHNICAL PART OF PATENTED INVENTION**19. Limitation of Prior Technology / Art**

NA

20. Specific Problem Solved / Objective of Invention

Method and system for detecting duplicate documents in web crawls

21. Brief about Invention

A Web crawler application takes advantage of a document store's ability to provide a content identifier (CID) having a value that is a unique function of the physical storage location of a data object or document, such as a Web page. In operation, the crawler first tries to fetch the CID for a document. If the CID attribute is not supported by the document store, the crawler fetches the document, filters it to obtain a hash function, and commits the document to an index if the hash function is not present in a history table. If the CID is available from the document store, the CID is fetched from the document store. The crawler then determines whether the CID is present in the history table, which indicates whether an identical copy of the document in question has already been indexed under a different URL. If the CID is present, indicating that the document has already been indexed, the new URL is placed in the history file but the document itself is not retrieved from the document store, nor is it filtered again to obtain a CID. If the CID is not present in the history table, the full document is retrieved and indexed. The CID data structure is an extension of a known globally unique ID (GUID). Whereas the GUID is a 16-byte number, the CID comprises a 16-byte GUID plus an additional 6-byte number.

22. Key learning Points

Web crawling

23. Summary of Invention

The present invention provides an improved way to access documents (including Web pages, file system documents, e-mail messages, etc.) stored in one or more document stores on a computer network. For example, the invention could be used in a Web crawler application, mail server, directory service, or any system requiring indexing or one-way replication of a document store. The invention is particularly directed to a method and system for identifying duplicate documents in a document store, and using this information to avoid unnecessarily retrieving and processing such duplicates.

24. Number of Claims : 22

25. Patent Status : Other (Active)

26. How much this invention is related with your IDP/UDP?

Not related to IDP/UDP, It's related to my area of interest

27. Do you have any idea to do anything around the said invention to improve it? (Give short note in not more than 500 words)

No. I don't have any idea about improvement on this. We have ideas of integrating this with other algorithms.



**GUJARAT TECHNOLOGICAL UNIVERSITY
(GTU)
INNOVATION COUNCIL (GIC)
Patent Search & Analysis Report
(PSAR)**



Date of Submission : 14/09/2017

Dear **Bhatt Jigar Maneshbhai**,

Studied Patent Number for generation of PSAR : 17BE7_150283116002_3

PART 1: PATENT SEARCH DATABASE USED

1. Patent Search Database used : Free Patents Online
Web link of database : <http://www.freepatentsonline.com/>
2. Keywords Used for Search : trend Ranking ,content Ranking,ranking
3. Search String Used : trend Ranking, content Ranking
4. Number of Results/Hits getting : 1016

PART 2: BASIC DATA OF PATENTED INVENTION /BIBLIOGRAPHIC DATA

5. Category/ Field of Invention :
6. Invention is Related to/Class of Invention : 1/1
6 (a) : IPC class of the studied patent : G06F17/30
7. Title of Invention : Content ranking based on user features in content
8. Patent No. :
9. Application Number : 14/147789
9 (a) : Web link of the studied patent : <http://www.freepatentsonline.com/9633119.html>
10. Date of Filing/Application (DD/MM/YYYY) : 01-06-2014
11. Priority Date (DD/MM/YYYY) :
12. Publication/Journal Number :
13. Publication Date (DD/MM/YYYY) :
14. First Filled Country : Albania

15. Also Published as

Sr.No	Country Where Filed	Application No./Patent No.
1		

16. Inventor/s Details.

Sr.No	Name of Inventor	Address/City/Country of Inventor
1	Wexler	Santa Carla
2	Mike	USA

17. Applicant/Assignee Details.

Sr.No	Name of Applicant/Assignee	Address/City/Country of Applicant
1	Yahoo Inc	Sunnyvale, CA, US

18. Applicant for Patent is : Individual

PART 3: TECHNICAL PART OF PATENTED INVENTION**19. Limitation of Prior Technology / Art**

NA

20. Specific Problem Solved / Objective of Invention

This pattern is about personalized news-reading experience . In order to provide a better news-reading experience

21. Brief about Invention

Methods, systems, and computer programs are presented for providing a personalized news stream to a user. One method includes an operation for identifying user features associated with a user. The user features include personal features and social features. The personal features are based on activities of the user and the profile of the user. The social features are based on information about social connections of the user. The method further includes operations for extracting content features from a corpus of content items, for identifying intersections between user features and content features, and for assigning weights to the content features from the corpus based on the identified intersections. A score for each content item is determined based on the content features and the respective weights of the content items. The content items are then ranked based on the scores. One or more of the ranked content items are displayed.

22. Key learning Points

Content ranking

23. Summary of Invention

The user features include personal features and social features, the personal features being based on activities of the user and based on a profile of the user, and the social features being based on information about social connections of the user. Further, the method includes operations for extracting content features from a plurality of content items, identifying intersections between user features and content features for the plurality of content items, and for assigning weights to the content features from the plurality of content items based on the identified intersections. Further, the method includes operations for determining scores for each content item based on the content features and respective weights of the content items, and for ranking the plurality of content items based on the scores. One or more of the ranked plurality of content items are then displayed for the user. In another embodiment, the operations of the method are executed by a processor.

24. Number of Claims : 20**25. Patent Status : Published Application****26. How much this invention is related with your IDP/UDP?**

Not related to IDP/UDP, It's related to my area of interest

27. Do you have any idea to do anything around the said invention to improve it? (Give short note in not more than 500 words)

No. I don't have any idea about improvement on this. We have ideas of integrating this with other algorithms.



**GUJARAT TECHNOLOGICAL UNIVERSITY
(GTU)
INNOVATION COUNCIL (GIC)
Patent Search & Analysis Report
(PSAR)**



Date of Submission : 11/09/2017

Dear Fotariya Jimish Maheshbhai,

Studied Patent Number for generation of PSAR : 17BE7_150283116008_4

PART 1: PATENT SEARCH DATABASE USED

1. Patent Search Database used : PatentScope (WIPO Patent Database)

Web link of database : <http://patentscope.wipo.int/search/en/search.jsf>

2. Keywords Used for Search : web crawling ,html parsing,content parsing

3. Search String Used : html parsing

4. Number of Results/Hits getting : 488

PART 2: BASIC DATA OF PATENTED INVENTION /BIBLIOGRAPHIC DATA

5. Category/ Field of Invention :

6. Invention is Related to/Class of Invention : 1/1

6 (a) : IPC class of the studied patent : G06F 17/30,G06N 99/00

7. Title of Invention : METHOD OF AND SYSTEM FOR CRAWLING A WEB RESOURCE

8. Patent No. :

9. Application Number : 15326045

9 (a) : Web link of the studied patent : <https://patentscope.wipo.int/search/en/detail.jsf?docId=US200949331&recNum=5&tab=Drawings&maxRec=488&office=&prevFilter=&sortOption=Pub+Date+Desc&queryString=FP%3A%28web+crawling%29>

10. Date of Filing/Application (DD/MM/YYYY) : 01/26/2015

11. Priority Date (DD/MM/YYYY) :

12. Publication/Journal Number :

13. Publication Date (DD/MM/YYYY) :

14. First Filed Country : Albania :

15. Also Published as

Sr.No	Country Where Filed	Application No./Patent No.
1		

16. Inventor/s Details.

Sr.No	Name of Inventor	Address/City/Country of Inventor
1	Damien Raymond JeanFrancois	USA
2	Liudmila Alexandrovna	USA
3	Egor Aleksandrovich	USA
4	Pavel Viktorovich	USA
5	Ivan Semeonovich	USA
6	Arsenii Andreevich	USA
7	Gleb Gennadievich	USA

17. Applicant/Assignee Details.

Sr.No	Name of Applicant/Assignee	Address/City/Country of Applicant
1	YANDEX	EUROPE AG

18. Applicant for Patent is : Group

PART 3: TECHNICAL PART OF PATENTED INVENTION**19. Limitation of Prior Technology / Art**

NA

20. Specific Problem Solved / Objective of Invention

method of setting up a crawling schedule

21. Brief about Invention

There is disclosed a method of setting up a crawling schedule, the method executable at a crawling server, the crawling server coupled to a communication network, the communication network having coupled thereto a first web resource server and a second web 5 resource server. The method comprises: appreciating a first new web page associated with the first web resource server; appreciating a second new web page associated with the second web resource server.

22. Key learning Points

Web crawling & its scheduling system

23. Summary of Invention

There is disclosed a method of setting up a crawling schedule, the method executable at a crawling server, the crawling server coupled to a communication network, the communication network having coupled thereto a first web resource server and a second web 5 resource server. The method comprises: appreciating a first new web page associated with the first web resource server; appreciating a second new web page associated with the second web resource server

24. Number of Claims : 22

25. Patent Status : Published Application

26. How much this invention is related with your IDP/UDP?

Not related to IDP/UDP, It's related to my area of interest

27. Do you have any idea to do anything around the said invention to improve it? (Give short note in not more than 500 words)

No. I don't have any idea about improvement on this. We have ideas of integrating this with other algorithms.



GUJARAT TECHNOLOGICAL UNIVERSITY
(GTU)
INNOVATION COUNCIL (GIC)
Patent Search & Analysis Report
(PSAR)



Date of Submission : 11/09/2017

Dear Fotariya Jimish Maheshbhai,

Studied Patent Number for generation of PSAR : 17BE7_150283116008_5

PART 1: PATENT SEARCH DATABASE USED

1. Patent Search Database used : Google Patents
Web link of database : <https://patents.google.com/>
2. Keywords Used for Search : Clustering on news ,Clustering on news ,Clustering on news
3. Search String Used : Clustering on news
4. Number of Results/Hits getting : 3668

PART 2: BASIC DATA OF PATENTED INVENTION /BIBLIOGRAPHIC DATA

5. Category/ Field of Invention :
6. Invention is Related to/Class of Invention : 1/1
- 6 (a) : IPC class of the studied patent : NA
7. Title of Invention : Methods and apparatus for clustering news content
8. Patent No. :
9. Application Number : US10611269
- 9 (a) : Web link of the studied patent : <https://patents.google.com/patent/US7568148B1/en?q=news&q=duplicate>
10. Date of Filing/Application (DD/MM/YYYY) : 06/30/2003
11. Priority Date (DD/MM/YYYY) :
12. Publication/Journal Number :
13. Publication Date (DD/MM/YYYY) :
14. First Filled Country : Albania

15. Also Published as

Sr.No	Country Where Filed	Application No./Patent No.
1		

16. Inventor/s Details.

Sr.No	Name of Inventor	Address/City/Country of Inventor
1	Krishna Bharat	NA
2	Michael Curtiss	NA
3	Michael Schmitt	NA

17. Applicant/Assignee Details.

Sr.No	Name of Applicant/Assignee	Address/City/Country of Applicant
1	Google Inc	Mountain View, CA

18. Applicant for Patent is : Group

PART 3: TECHNICAL PART OF PATENTED INVENTION**19. Limitation of Prior Technology / Art**

NA

20. Specific Problem Solved / Objective of Invention

clustering news content

21. Brief about Invention

Methods and apparatus are described for scoring documents in response, in part, to parameters related to the document, source, and/or cluster score. Methods and apparatus are also described for scoring a cluster in response, in part, to parameters related to documents within the cluster and/or sources corresponding to the documents within the cluster. In one embodiment, the invention may identify the source; detect a plurality of documents published by the source; analyze the plurality of documents with respect to at least one parameter; and determine a source score for the source in response, in part, to the parameter. In another embodiment, the invention may identify a topic; identify a plurality of clusters in response to the topic; analyze at least one parameter corresponding to each of the plurality of clusters; and calculate a cluster score for each of the plurality of clusters in response, in part, to the parameter.

22. Key learning Points

Items clustering

23. Summary of Invention

"Methods and apparatus are described for scoring documents in response, in part, to parameters related to the document, source, and/or cluster score. Methods and apparatus are also described for scoring a cluster in response, in part, to parameters related to documents within the cluster and/or sources corresponding to the documents within the cluster. In one embodiment, the invention may identify the source; detect a plurality of documents published by the source; analyze the plurality of documents with respect to at least one parameter; and determine a source score for the source in response, in part, to the parameter. In another embodiment, the invention may identify a topic; identify a plurality of clusters in response to the topic; analyze at least one parameter corresponding to each of the plurality of clusters; and calculate a cluster score for each of the plurality of clusters in response, in part, to the parameter.

Additional aspects of the present invention are directed to computer systems and to computer-readable media having features relating to the foregoing aspects."

24. Number of Claims : 28

25. Patent Status : Published Application

26. How much this invention is related with your IDP/UDP?

Not related to IDP/UDP, It's related to my area of interest

27. Do you have any idea to do anything around the said invention to improve it? (Give short note in not more than 500 words)

No. I don't have any idea about improvement on this. We have ideas of integrating this with other algorithms.

Conclusion and flow

So how all things will proceed?

Before fetching data, we will need to have the targets to get data from. For that we will be gone list out most popular news feeders around the world.

There will be three main phases of the flow

- Crawling content from Internet and load to the data center.
- Parsing and indexing content and compute relevancy.
- Publishing content on the portal server.

Finding information by crawling.

We use software known as “web crawlers” to discover publicly available web pages. The most well-known crawler is called “TIMESWENBOT.” Crawlers look at web pages and follow links on those pages, much like you would if you were browsing content on the web. They go from link to link and bring data about those web pages back to TIMESWEN’s servers.

The crawl process begins with a list of web addresses from past crawls and site-maps provided by website owners. As our crawlers visit these websites, they look for links for other pages to visit. The software pays special attention to news sites, changes to existing sites and dead links.

Computer programs determine which sites to crawl, how often, and how many pages to fetch from each site. TIMESWEN doesn't accept payment to crawl a site more frequently for our web search results. We care more about having the best possible results because in the long run that's what best for users and, therefore, our business.

We crawl under restriction policies like robot.txt. Most websites don't need to set up restrictions for crawling, indexing or serving, so their pages are eligible to appear in search results without having to do any extra work. That said, site owners have many choices about how TIMESWEN crawls and indexes their sites through Webmaster Tools and a file called “robots.txt”. With the robots.txt file, site owners can choose not to be crawled by bot, or they can provide more specific instructions about how to process pages on their sites.

Organizing information by indexing.

The web is like an ever-growing public library with billions of books and no central filing system. TIMESWEN essentially gathers the pages during the crawl process and then creates an index, so we know exactly how to look things up. Much like the index in the back of a book, the TIMESWEN index includes information about words and their locations. When you search, at the most basic level, our algorithms look up your search terms in the index to find the appropriate pages.

The search process gets much more complex from there. When you search for “dogs” you don’t want a page with the word “dogs” on it hundreds of times. You probably want pictures, videos or a list of breeds. TIMESWEN’s indexing systems note many different aspects of pages, such as when they were published, whether they contain pictures and videos, and much more. With the Knowledge Graph, we’re continuing to go beyond keyword matching to better understand the people, places and things you care about.

We focus that you always get the Relevant content, that’s thy parsing is very important phase. After content will get parsed, then and then we can compute the relevancy of content according the algorithm that works like Facebook's news feed algorithm known as Edge-Rank.