

Designing and Implementing Efficient Web-based RSS Reader for Mobile Devices

Priya Kaul

Department of Computer Engineering
K.J. Somaiya College of Engineering
University of Mumbai, India

Deepak Sharma

Department of Computer Engineering
K.J. Somaiya College of Engineering
University of Mumbai, India

ABSTRACT

RSS (Really Simple Syndication) is a content delivery vehicle. It is the format used when we want to syndicate news and other web content. When this content gets distributed it is called a feed. It takes most people much time to look for new information from one website to other. With RSS readers, we can aggregate latest updates from multiple websites in one place creating a single view for this information. The application explained implements a web-based RSS reader which incorporates an Automatic Filtering Technique (AFT) for mobile devices. A web-based RSS reader is implemented that will simply aggregate news video feeds from multiple news sources. After that, AFT (Automatic Filtering Technique) is to be implemented which forms the crux of this project. If a mobile user is interested in receiving a specific video in reference with a specific category, he or she need not scroll through the whole RSS feeds- which is the usual scenario. In fact, user can submit request as per his requirements, and receive the appropriate video via RSS reader. Repeated requests from same user are retrieved from a buffer and not from the web server. This filtering technique will surely help in reducing the processing cost of mobile devices. Moreover, system is developed as web-based RSS reader solution so as to allow the users to retrieve updated content through the website by only using a mobile web browser. They need not install a mobile reader on device thereby saving resources.

General Terms

RSS reader, processing cost

Keywords

RSS (Really Simple Syndication), AFT (Automatic Filtering Technique)

1. INTRODUCTION

In the era of web 2.0, everyone can create and update content, and everyone can host a personal web site with little effort, making it hard to gather valuable information from different web sites. There is too much news out there nowadays for anybody to keep up with the pace of information update. Individuals might look to different portals for news, stocks, security warnings, press releases, industry analysis, product reviews, and so on. Traditionally, this process was time consuming for users as they visited each Web site, poking around for new information before moving on to another website. RSS technology has paved way for quick retrieval of updates from multiple websites at the same time on the user's device. Generally mobile devices and for the matter of fact desktop systems also need to install these RSS readers, which is followed by subscription to different sites. Need is to

develop an online system that not only provides basic functionality of RSS readers, but also help in improving overall experience of users by giving them freedom to choose the search criteria for their requests. Video feeds is a derivative of Text feeds only, but not much work has been done to make them available to the users as easily as text feeds. Section 2 explains evolution of RSS technology through various stages. Right from the push/pull technology to the concept of making RSS readers intelligent, everything is discussed in detail. Section 3 explains the general architecture of system to be implemented -efficient web based RSS reader for mobile devices. Further subsections explain development of subsystems namely Proxy server, news website and Automatic Filtering Technique in detail. Section 4 explains analysis of results of the system. Website is made to run on local host and proxy server is executed using WCF (Windows Communication Foundation) client. Section 5 concludes the whole project by summarizing the key points, advantages, and edge factors that make this application worth usable and useful for all smartphone users who want to have access to the latest, shortest and most seen news videos from all corners of world without actually navigating from one website to other and thus minimizing browsing time.

2. REVIEW OF LITERATURE

RSS providers such as news sites or weblogs, need to provide the contents as XML document called RSS feed which can be displayed using a reader or aggregator. Users provide RSS feed URL to the reader, which will feed the contents to be displayed on a user's machine, and will automatically check for updated contents. The contents are only concise textual contents like news topics or news summaries. The whole content can be viewed by clicking on the relevant link. In 1997, RSS technology [1] was developed to aggregate updated contents according to push technology. This paper explains protocols used by information feed i.e. Internet protocols such as HTTP and TCP/IP. MCF (Microsoft Connector Framework) and CDF (Channel Definition Framework) are the precursors to RSS. These are the technologies that paved way for development of RSS. RSS has several versions which are still in widespread use. RSS 0.91 remains a frequently used information feed technology among the feeds listed on Syndic8.com. RSS 0.91, 1.0, and 2.0 and Atom 0.3 all have widespread usage. RSS 0.92 has significant usage but perhaps an order of magnitude less than the other four technologies just mentioned. Technology to aggregate contents on the web originated from "push" and "pull" technology [2]. Push services are often based on information preferences expressed in advance. This is called a publish/subscribe model. A client might subscribe to various

information "channels". Whenever new content is available on one of those channels, the server would push that information out to the user. For pull technology, the initial request for data comes from the client, while push technology requests for a given transaction created by a publisher or central server. The server will specify what and when the content will be shown. Aldrich, Bell and Batzel [3] developed a system that provides a solution which automatically generates the podcasts of class lectures and publishes them on online class blog-space. The system requires class information to record the audio at a specific time, and creates the RSS feed. A link to a finished audio file will then be added to the feed. Curran et al. Paper [4] proposes a Scheduled RSS Delivery System, which utilizes idle computer time to subscribe to media RSS channels for downloading podcasts. The aggregator in the system will scan for newly added items and place those on the download list when the computer is idle. Subsequently, this system was improved so that the class lectures can be recorded as podcast in both audio and video format. Blekas et al. [5] proposed a method to adapt the web content before it can be accessed by a mobile browser due to its technical limitations. They built a proxy server solution that is online system using RSS feeds for an adaptation of the web content for mobile phones. The existence of RSS feeds significantly improves the content presented to the mobile users, decreases both size and access cost because RSS feeds contain metadata that summarize information. The proxy server, which usually runs on powerful server, unleases mobile devices from computational needs. However, the limitations of this method are that images and tables will be eliminated from the content and HTML forms cannot be used. Paper [6] proposes and implements the OWeB system- a framework for Offline Web Browsing- to improve the Internet browsing experience over slow and intermittent networks. The aggregator of the system needs to download the RSS feed to construct the homepage. Thus the results in websites are up-to-date and available offline, and also help in bandwidth savings. With RSS, people can read information from different resources in a uniform way, and in a single tool, such as an RSS reader. However, most of the RSS readers only display items in chronological order, which doesn't work well when users are inundated with too many items in the feeds. Various recommendation features [7] have been proposed like profiled based features (i.e., text similarity and favorite fraction), update frequency, as well as PostRank values for RSS recommendation. Experimental results have indicated that favorite fraction and update frequency perform better than text similarity. Effect of feature combination is also studied and it is found out that the combination of similarity and favorite fraction performs the best.

Traditional recommendation systems do the data mining on web access logs, discover user's access patterns, and filter the information on behalf of the user at the server side. One critical limitation of traditional recommendation system is the lack of user's private daily data, such as schedules, favorite websites and personal emails. The reason for this limitation is the privacy leak issue when the server holds much more private user data. To solve this problem, this paper presents an agent-based personalized recommendation method called Content Recommendation System based on private Dynamic User Profile (CRES DUP) [8]. The system collects and mines the private data of user at the client side, discovers, stores and updates private Dynamic User Profile (DUP) at the client side. The system fetches preferred message from the content server according to DUP. An important usage of this technology is a personalized advertising system in the RSS (Rich Site

Summary, or RDF Site Summary) reader application. [9] This paper explores the effect that blogs (the end product) have had on the adoption of a related technology, RSS, by online newspapers. It uses the diffusion of innovation models presented in Everett Rogers' *The Diffusion of Innovation* and Brian Winston's *Media, Technology and Society* to explain the relationship between RSS, blogs, and online news. The paper first explores the origins of RSS technology before moving to its adoption and diffusion. It then outlines adoption and barriers to adoption in the online news sector.

Collaborative Filtering (CF) [10] has achieved widespread success in recommender systems such as Amazon and Yahoo! music. However, CF usually suffers from two fundamental problems - data sparsity and limited scalability. Among the two broad classes of CF approaches, namely, memory-based and model-based, the former usually falls short of the system scalability demands, because these approaches predict user preferences over the entire item-user matrix. The latter often achieves unsatisfactory accuracy, because they cannot capture precisely the diversity in user rating styles. In this paper, we propose an efficient Collaborative Filtering approach using Smoothing and Fusing (CFSF) strategies. CFSF formulates the CF problem as a local prediction problem by mapping it from the entire large-scale item-user matrix to a locally reduced item-user matrix. Given an active item and a user, CFSF dynamically constructs a local item-user matrix as the basis of prediction. To alleviate data sparsity, CFSF presents a fusion strategy for the local item-user matrix that fuses ratings of the same user makes on similar items, and ratings of like-minded users make on the same and similar items. To eliminate diversity in user rating styles, CFSF uses a smoothing strategy that clusters users over the entire item-user matrix and then smoothes ratings within each user cluster. Empirical study shows that CFSF outperforms the state-of-the-art CF approaches in terms of both accuracy and scalability.

3. EFFICIENT WEB BASED RSS READER FOR MOBILE DEVICES

System to be implemented consists of three elements: Proxy server, Web server and a data repository. A Proxy Server (PS) checks if the request from the user is similar to the previous request. If so, the filtered video is sent to the user's device. If request submitted by the user is new i.e. not a repetition of previous request, proxy server forwards request to Web Servers (WS) of three different news websites. These websites are- msn, Times and NDTV. Web servers of websites reply with the video feeds. As soon as proxy server receives the replies, it updates the results in data repository. Data repository acts as a buffer that is used to hold the retrieved video feeds. General architecture of the system is explained diagrammatically in figure I.

3.1 Proxy server development

Proxy server is developed using WCF (Windows Communication Foundation). There are a number of existing approaches to building distributed applications. These include Web services, .NET Remoting, Message Queuing (MSMQ) and COM+/Enterprise Services. Windows Communication Foundation (WCF) unifies these into a single framework for building and consuming services. Microsoft originally introduced WCF as part of the .NET Framework 3.0 and has continued to enhance it for the .NET Framework 3.5 and Visual Studio 2008. WCF has an edge over other technologies

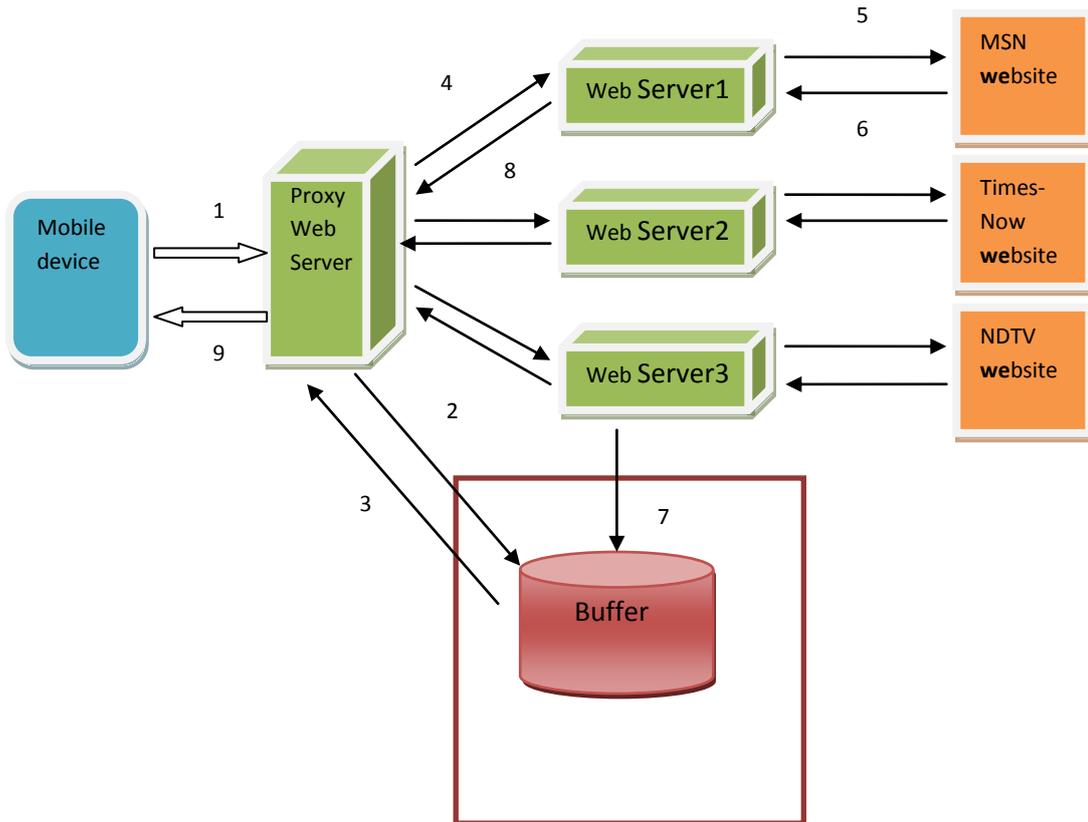


Fig I- Web based RSS reader for mobile devices

like Web services and .Net Remoting. It is the proxy server that accepts requests from the news website on the behalf of user. Proxy server fires the request on the three news web servers namely- msn,Times and NDTV. The arrangement of video rss feeds on the server varies from site to site. No two sites employ same organization pattern for their respective RSS feeds. Web scraping (web harvesting or web data extraction) is a computer software technique of extracting information from websites. The webpages containing RSS feeds contain thousands of lines written in HTML. Tremendous effort has been done in the web scraping phase. Only necessary data that contains the video link and their duration is selected from the web pages containing video rss feeds. For scrapping of required data, concept of regular expressions is implemented. "System.Text.RegularExpressions" is the header file that supports functions provided by "Regex" class. Using HttpWebRequest we are able to scrap the useful data and remove unwanted data from HTML pages. Proxy server doesn't have any User interface. It simply provides services to the front end application. Since website is not developed in the first phase, we can see the results of execution of proxy server in WCF test client window.

3.2 Data Repository

News database contains three tables namely- categories, date and links. Data repository acts in close coordination with proxy server. Structure of news database is explained in figure II.

3.3 News website

Microsoft Visual Studio is the technology used to develop the front end of application i.e. news website. ASP.NET is a server-side web application framework designed for Web development to produce dynamic Web pages. News website provides user with four categories namely News, Entertainment, Sports and Business. Also user is given an option for refining his search criteria by opting for any of the four options – Duration, Title, Reviews, Best video.

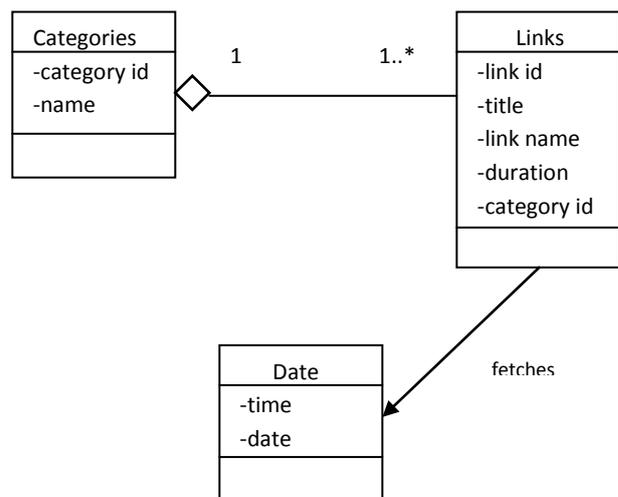


Fig II. News database structure

3.4 Automatic Filtering Technique

Automatic Filtering Technique implements two important aspects of the system. First aspect enhances the efficiency of system in retrieving video feeds from the websites. And second aspect helps in saving the resources of mobile devices by fetching results directly from database and note from the web server. First aspect exploits web scrapping capability of proxy server. Since each of the websites have different organizational pattern for their video RSS feeds, regular expressions for scrapping the useful part vary from site to site. The HTML format of web page displaying video RSS feeds has nearly thousands of lines in it. Hence great amount of accuracy was needed in designing the regular expressions. First aspect is represented diagrammatically in figure III. In second aspect of AFT- a duration criteria for checking whether request submitted by a user is similar to his previous request is set. As per the requirement this criteria can be changed. For experimental purpose duration criteria is set to 5 hours. If time difference between users's current and previous request is less than 5 hours, resultant video feeds are retrieved from buffer i.e. news database instead of hitting web servers. This not only helps in fetching faster results but also helps in saving mobile resources.

```
If (request is new)
    for each new request (criteria entered by the user
        is one out of duration, title, reviews, best video)
        download feed content from the webservers
            for each item in the feed
                Get data in <title>,<duration> and
                <category name> format and forward to proxy
                server
                proxy server forwards the results to buffer
            end for
        end for
        Retrieve results from proxy
Else
    Retrieve result from buffer and show to user
```

Fig III First aspect of Automatic Filtering Technique

This first aspect is better explained with the help of fig I under following points (numbering of the points is as per the diagram, not to be considered as bullets)-

- 1- user submits the category from four options: News, Entertainment., Sports, Business. After that one of the four search criteria is entered, namely- Duration, Title, Review, Best video. Request is forwarded to proxy server.
- 2- request is forwarded to buffer, and there it is checked if same request has been made by the user in past 5 hours, if its not so, then following steps are carried out.
- 3- request is sent back to proxy server
- 4- proxy server forwards the request to web servers.
- 5- web servers collect video feeds from websites
- 6-7, 6-8 Results from web servers are forwarded to buffer. And at the same time results are also forwarded to proxy server

8-9 Proxy server forward replies to the user's mobile device

Second aspect of AFT is explained with the help of fig I under following points-

- 1- user submits the category from four options: News, Entertainment., Sports, Business. After that one of the four search criteria is entered, namely- Duration, Title, Review, Best video. Request is forwarded to proxy server.
- 2- request is forwarded to buffer, and there it is checked if same request has been made by the user in past 5 hours, if yes then instead of hitting web server, results are directly fetched from buffer
- 3- results received by proxy server
- 9- proxy server forwards final result on user's mobile device.

4. EXPERIMENTAL RESULTS

Front end of the application i.e. news website is developed using ASP.Net. Proxy server development is carried out using WCF technology. Database used is Microsoft SQL Server. Application is executed using IIS manager. Results are summarized keeping two aspects of AFT in mind. In first aspect, application allows user to submit requests by letting him select one out of four categories as per their requirement. These are- News, Entertainment, Sports, Business. After selecting one of these options, user can further refine his search by selecting one out of four search criteria- duration, title, reviews, best video. If user selects duration option, first five video feeds in ascending order of their duration are displayed on the user's mobile device. All these feeds have been collected from three websites- msn, TimesNow and NDTV. If user selects title option, he ought to enter any keyword of his interest but that keyword should be part of any ongoing news event then only first five feeds containing that keyword in their title field will be displayed on user's device. If option selected is reviews- then top five feeds in the ascending order of their reviews are displayed on user's device. Last option is Best video option. In this option, user get to see top five feeds on his mobile screen which satisfy the criteria that- number of reviews should be greater than zero and duration of video should be less than three. All these results satisfy the first aspect of AFT. Second aspect of AFT focuses on fast retrieval of results and saving processing power of mobile devices. As per this part of AFT, if a user submits a request which is similar to any request made by him in past five hours, then instead of getting results by hitting proxy server, results are fetched directly from buffer i.e. news database.

5. CONCLUSION

By introducing four search criteria duration, title, reviews and best video, user's experience of application is enhanced. Especially title option eases out user's burden of browsing through all video feeds from a specific category and looking for the feed of his choice. Second aspect of AFT i.e. fetching results directly from buffer without hitting web servers of news websites greatly enhances the efficiency of system. Processing power of device along with battery consumption is saved. Mobile devices need not install any application, they can directly access website via mobile browser.

6. FUTURE RESEARCH

Filtering technique used in this project is limited to only few criteria. Techniques like collaborative filtering and content recommendation can be incorporated in the project so as to make system more intelligent. Fusion of online system with artificial intelligence will considerably enhance the usefulness of system. Also efforts will be made to increase the number of categories so that user base of the website is increased considerably. More news websites can be added as the video feed providers so that users get access to huge variety of feeds from varying sites.

7. REFERENCES

- [1] A. Watt and D. Ayers, "Beginning RSS and ATOM Programming", Wiley Publishing, Inc., Indianapolis, Indiana, 2005.
- [2] M. Franklin, and S. Zdonik, "'Data in your face': push technology in perspective". Proceedings of ACM SIGMOD international conference on Management of data, New York, USA, 1998, pp. 516-519.
- [3] D. Aldrich, B. Bell, and T. Batzel, "Automated podcasting solution expands the boundaries of the classroom". Proceedings of the 34th annual ACM SIGUCCS conference on User services, New York, USA, 2006, pp. 1-4.
- [4] K. Curran, S. Mc Kinney, F. Burns, and G. Meredith, "Scheduled RSS feeds for streaming multimedia to the desktop using RSS enclosures", Consumer Communications and Networking Conference, 2006, CCNC. 3rd IEEE, pp. 267-271.
- [5] A. Blekas, V. Garofalakis, and V. Stefanid, "Use of RSS feeds for content adaptation in mobile web browsing", Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A), New York, USA, 2006, pp. 79-85.
- [6] G. Ananthanarayanan, S. Blagsvedt, and K. Toyama, "OWEB: a framework for offline web browsing". Web Congress LA-Web Fourth Latin American, 2006, pp. 15-24.
- [7] Cansheng Ji, Jingyu Zhou, "A Study on Recommendation Features for an RSS Reader", International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2010
- [8] Ting Chen, Wei-i Han, Hai-Dong Wang, Ti-Xun Zhou, Bin Xu, Bin-Yu Zang, "Content recommendation system based on private dynamic user profile", Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007
- [9] Kathy E. Gill, "Blogging, RSS and the Information Landscape: A Look At Online News", 2005
- [10] Daqiang Zhang, Jiannong Cao, Jingyu Zhou, Minyi Guo, Vaskar Raychoudhury, "An Efficient Collaborative Filtering Approach Using Smoothing and Fusing", 2009