Use of Yahoo! Pipes to deliver information to farmers

Anita Gutta[1] and Dr. Priti Srinivas Sajja [2]

[1, 2] Department of Computer Science & Technology
Sardar Patel University, Vallabh Vidyanagar, India

ABSTRACT

The term "Information overload" was first used by futurologist Alvin Toffler in 1970. Over the years the problem has been increasing exponentially. With the ease in the methods in which information can now be created and passed there is no easy end to this problem. Only those who can deal with it can take advantage of all the available knowledge from different sources. Most of the farmers are not web or technology savvy to navigate them through the sea of information. This paper proposes how Yahoo! Pipes can be used to aggregate information from different sources and apply filtering modules on the output to deliver just the right content to farmers.

Keywords: Information overload, Information Retrieval system, Yahoo! Pipes.

1. INTRODUCTION

Information on the web is growing exponentially. The information is not just on the web but different mediums of information sharing have evolved over the years. The data formats are different and plain search engine browsing for data may not get to the right information sources. The ease with which people are able to create data has also added complexity to the problem. With multiple people keying their input it is hard to gauge the integrity of the data. This is true for all fields and is even more significant in farming sector. The availability and reliability of internet and other information delivery sources are at a minimum. If the situation in India is taken into account the literacy rate among farmers is very low so even if the infrastructure is provided they may be lacking the skills to use them properly. And with the "information overload" problem they have to be extremely crafty in how they frame questions also.

2. PROPOSAL

Getting data from multiple sources is not a new concept. After the advent of internet and as the web started growing exponentially there are lot of 3rd party tools and web crawlers which browse the web in a methodical and orderly manner. Even though these work to a certain extent there is a lot of code that needs to be maintained periodically. These tools also get the whole sites or the contents on a whole page and there is not much intelligence that can be added to get just the preferred sections. So as seen these tools help gather the content in one place there is still a lot of refinement and massaging that can be and needs to be done on the results.

This is where Yahoo! Pipes come in handy. They provide a very easy and convenient way to selectively aggregate data to meet personal needs. It is a hosted service provided by Yahoo that lets you remix feeds and create new data mashups in a visual programming environment. In the next section we will go over pipes implementation details.
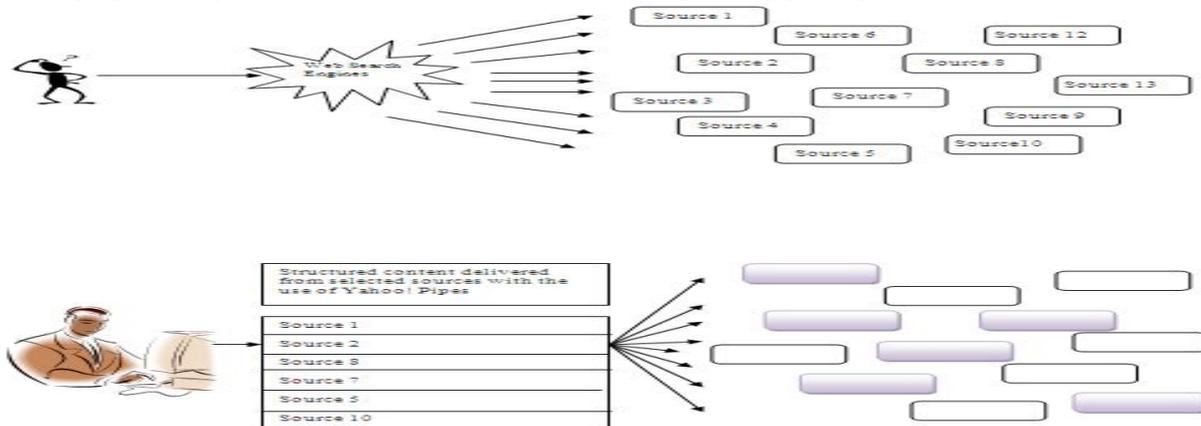

Figure 1 Information gathering from WEB and with the use of Yahoo! Pipes

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org, editorijaiem@gmail.com
Volume 2, Issue 1, January 2013 **ISSN 2319 - 4847**

3. IMPLEMENTATION

Yahoo! Pipes is a web-based, visual programming environment introduced by Yahoo! in 2007 with the intent of enabling users to "rewire the web". Pipes is a dataflow system in which the data is sourced from the web (RSS feeds, web pages, raw data) and flows through an interconnected set of modules that act upon it, ultimately producing some result; the Pipes name is inspired by the concept of pipes in Unix operating systems that enabled the composition of command line sequences of Unix tools through which data "flow" for processing. [3]

Yahoo Pipes is a particularly interesting utility that sits between an application and a source of data. This service allows developers to transform, fine-tune, and otherwise experiment with content before pulling it into an application. It is a simple but powerful mechanism that saves developers much work when they build an internal parsing and editing engine within their applications [4].

Creating data feeds from pipes is very easy. With a few simple steps data can be aggregated. The tool is extremely user-friendly so people with less computer experience can easily navigate through the steps. Advanced filtering techniques can be applied to further narrow down the results. Each new pipe link created can be stored and published. Other users can browse through publicly available pipes before creating a new one.
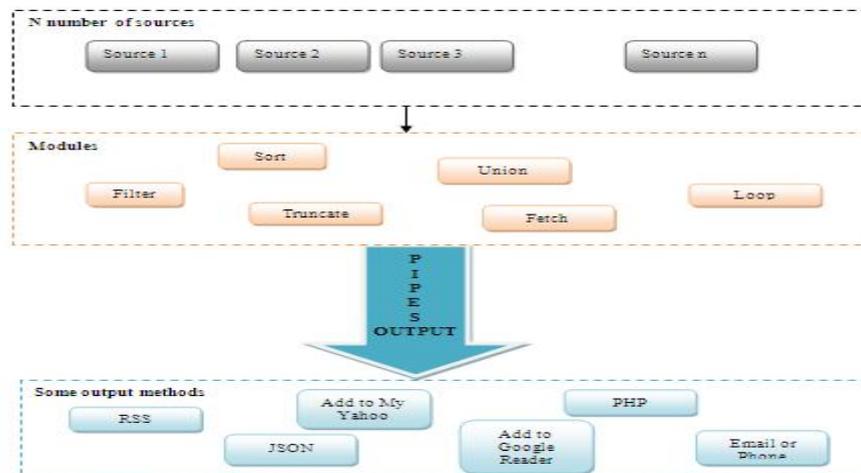
Figure 2 Process of building Pipes

In the sample pipe shown below data from two separate URL's are aggregated into a single data source. The results can be seen at the bottom of the screen.
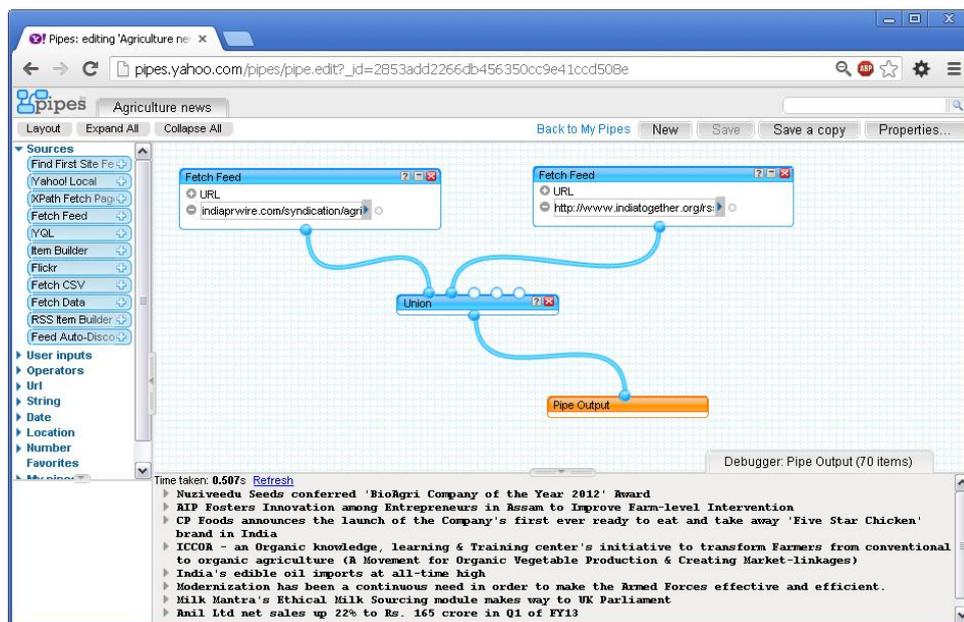

Figure 3 Demonstration of a simple Pipe

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org, editorijaiem@gmail.com
Volume 2, Issue 1, January 2013   **ISSN 2319 - 4847**

When the pipe is run it produces the results in a list like below.

Figure 4 Sample output of a simple pipe

The pipe designed earlier can be refined by adding filtering modules. From the first site only items whose description matches "seeds" are selected. Similarly from the second site any items which have description of "rice, wheat, sugarcane, barley or cotton" are selected. Data can be permitted or blocked as needed. The results from these are sent to a union module. Then results are sent to Sort module and results are sorted in ascending order. Truncate module limits the number of results returned. Only a few options are shown here more complex rules can be built depending on the output desired.
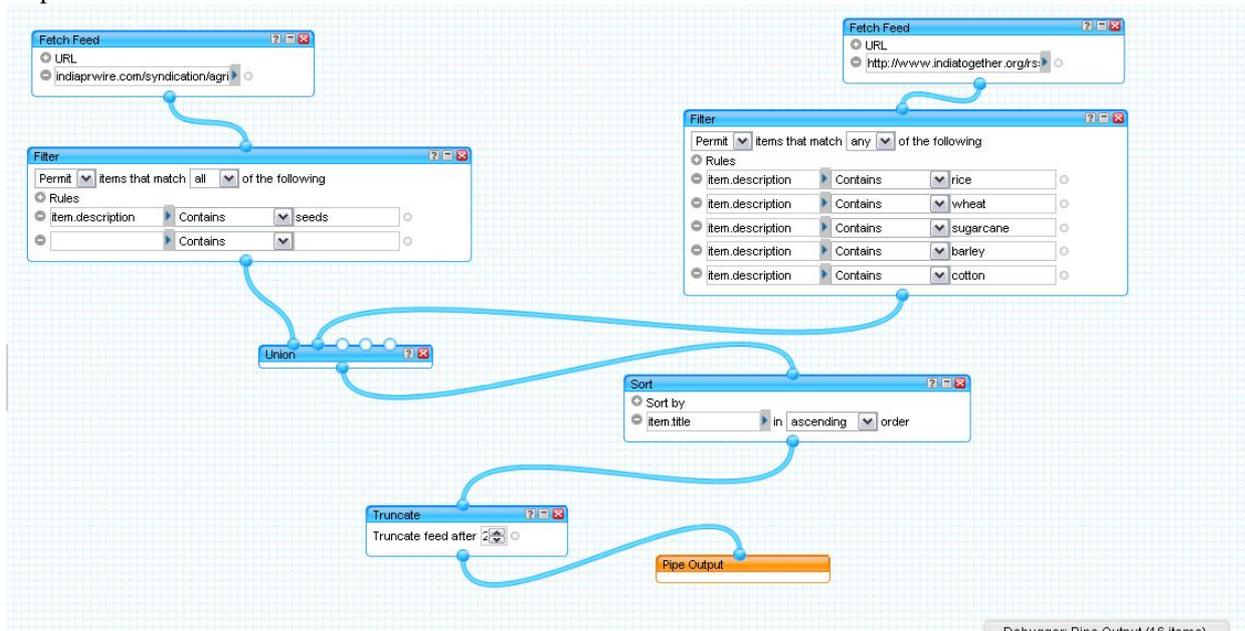

Figure 5 Sample of a Pipe with more modules

When the above pipe is run it gives the data as below. The output is limited to the conditions given in filter modules and are sorted alphabetically. Truncate module limits the output to top 16 matches.

Figure 6 Simplified output with the addition of filtering modules in Pipes

As shown above with all the different modules available on Yahoo! Pipes it is easier to create a very personalized content to users. As we have mentioned before literacy rate among farmers in India is very low and with the different languages spoken in different parts of the country pipes will be very handy to construct content in different language. These links can be saved in database and based on users region content will be presented in their own language. This kind of personalization will be very hard if done by programming methods. As these pipes are so easy to create users can also be given an option to create their own pipes and save it in their profile so they don't have to be bound by admin generated links.

Very few of the features are covered in this paper. Below is how a more well used and complicated pipes implementation looks like [7].
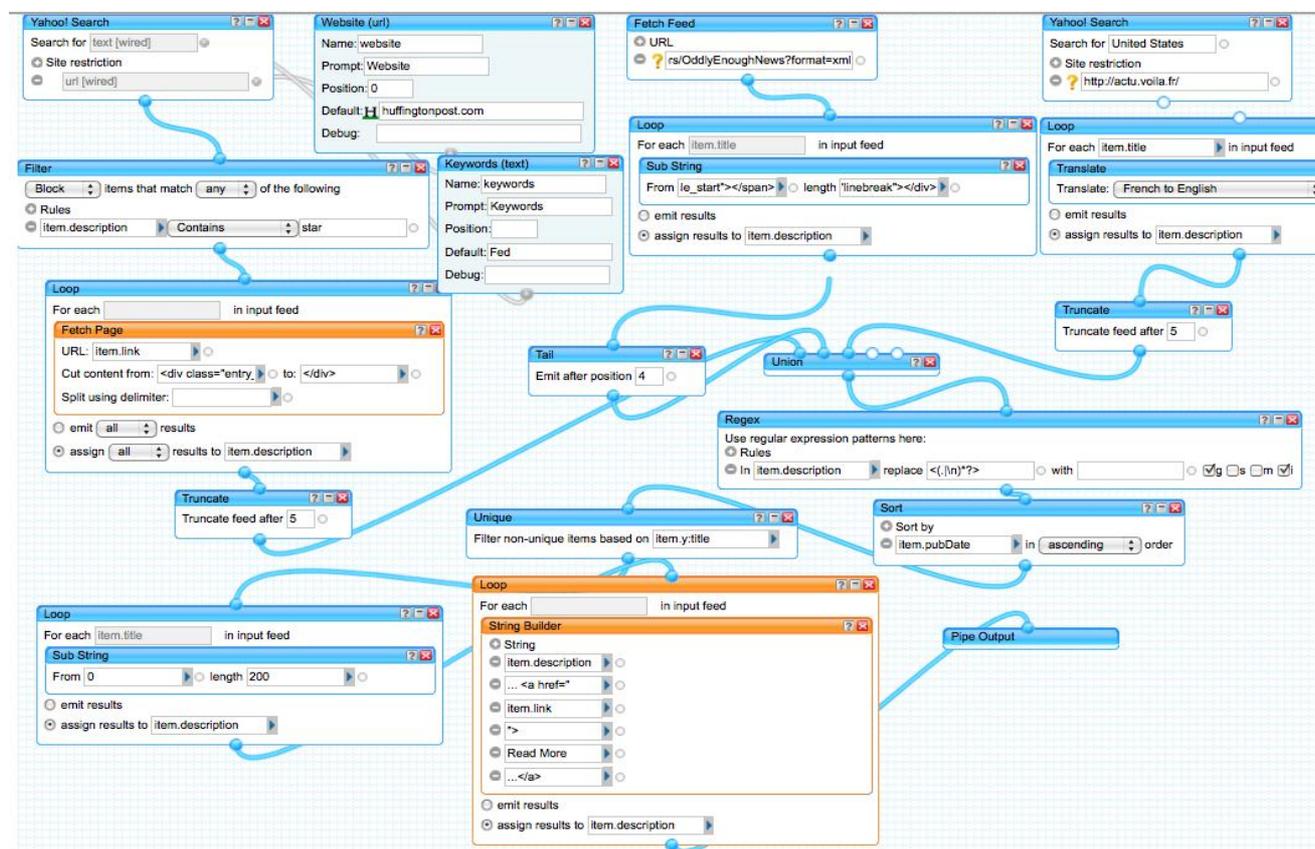

Figure 7 A more complicated design on Pipe

4. CONCLUSION

The importance of aggregating data from different source is seen from a long time and as shown in this paper this process is greatly simplified and made easier with the use of Yahoo! Pipes. Yahoo Pipes is a powerful composition tool to aggregate, manipulate, and mashup content from around the web. You can grab a collection of your favorite RSS feeds and manipulate them in Yahoo Pipes. This tool is not just used to get data from a few sites or keep track of different data on various social networking sites it is also now used by corporations to monitor their competitor sites. Retailers have been using it to collect intelligence about industry trends and to come up with more competitive pricing strategies

REFERENCES

[1] http://pipes.yahoo.com/pipes. [Online]
[2] http://www.infogineering.net/understanding-information-overload.htm. [Online]
[3] Empirically-Observed End-User Programming Behaviors in Yahoo! Pipes. http://www.ppig.org/papers/22nd-EUP-3.pdf.[Online]
[4] http://www.ibm.com/developerworks/xml/tutorials/x-yahoopipes/section2.html. [Online]
[5] Information Overload, Retrieval Strategies and Internet User Empowerment. Carlson, C.N. s.l. Proceedings of The Good, the Bad and the Irrelevant, 2004, Vols. Information overload, retrieval strategies and Internet user empowerment.
[6] The problem of information overload in business organisations: a review of the literature. Angela Edmunds, Anne Morris. 1, s.l. : International Journal of Information Management, 2000, Vol. 20.
[7] http://www.newthinktank.com/2010/08/yahoo-pipes-tutorial-part-3/yahoo-pipes-4-2/ [Online]

AUTHOR

Mrs. Anita Gutta received her M.S. degree from University of Missouri Kansas City in 2000. Currently pursuing her PhD from Sardar Patel University, Gujarat, India. Main research areas include Multiagent systems, Ontology and Fuzzy Logic