# Vision

## Introduction

News Feeder is a news feed aggregator service, which compiles content from multiple online content sources into a common news sheet format. It is intended for use by anyone in the general public that wishes to view regularly updating content in a simple manner.

### Glossary

**Item/Article –** An individual piece of content. This includes articles from RSS feeds or items from other types of content sources (such as an email, forum thread, weather forecast, etc).

**Feed –** A collection of items/articles all related to the same content source (i.e. the same site or RSS feed).

**Sheet –** A grouping method for feeds. It displays feeds on the website as views the user has created.

**Sheet Layout –** The view used to display feeds and content associated with the sheet. An individual sheet may have many of these views. When referring to a sheet and the way it displays its content, we are implicitly referring to its default view.

**Front-end –** The web interface of the system, where users gain access to and interact with the system.

**Back-end –** An application which runs a continuous queue and crawls/parses content sources for articles to add to the database.

### Problem Statement

|  |  |
| --- | --- |
| **The problem of** | Retrieving and viewing content from multiple online content sources that are updated regularly |
| **affects** | Any member of the general public who wishes to stay up to date with all this information; |
| **the impact of which is** | as they have to visit each location for its relative content which is time consuming and it is easy to miss information, particularly for regularly updated content. |
| **A successful solution would be** | A service which crawls content sources specified by a user and retrieves content to display in a practical and suitable format. Meanwhile, requiring as little interaction/configuration from the user as possible to continue retrieving this content. |

### Position Statement

|  |  |
| --- | --- |
| **For** | Any members of the general public |
| **who** | Wish to view regularly updated online content (such as RSS feeds) without needing to manually visit each content source. |
| **Newsfeeder is a** | Online news aggregation service |
| **that** | compiles content from multiple online content sources into a common news sheet format. |
| **Unlike** | the alternative of accessing each content source for its relative news manually, and at regular intervals (to ensure no news is missed), |
| **our product** | Automatically crawls the content source and retrieves new content as it becomes available, and displays it in an attractive, organised and customisable format. This ensures no content is missed and it is easily accessible for the user via an online interface. |

## System Overview

News Feeder is a news feed aggregation service, which compiles content from multiple online content sources. Its main purpose is to provide a single location and format to view news content instead of users having to access multiple websites for their specific news.

News Feeder allows users to specify which type of articles they wish to view, based on a number of criteria, including; location, category, author, publishing site and specific feed URLs.

The system crawls feeds which match the specified criteria and stores the relevant articles/items and their content. The articles/items can come from a number of different content sources (such as RSS, forums and social networks) and their content can be of any media type, including; text, images, video and geo-locational information.

The user can then view the stored articles and the article content on their sheets. The layout of the sheet will be in a standard news sheet format by default, but the system also provides a simple drag and drop interface for creating custom user layouts for sheets. Sheets will allow users to set filtering options to ensure certain types of content are not displayed to inappropriate audiences (such as pornographic material to minors).

Users can choose to subscribe to the News Feeder service for a monthly fee. This provides them with a number of benefits over standard users, such as access to the downloader application, customisable sheet layouts, increased feed/sheet limits and increased priority for content retrieval.

The system provides access for administrators to manage the News Feeder service. Administrators can manage users, feeds and items, as well as managing and monitoring back-end system processes (such as starting and stopping the crawler).

Access to this functionality is available via a web-interface. Users may also access News Feeder content offline via the use of a downloader application.

### Assumptions and Dependencies

The following assumptions and dependencies relate to the development and capabilities of the Newsfeeder service:

* The Newsfeeder system is intended to provide a service to the general public for use. It is not a product marketable to third-parties for implementation. This in turn adds the following assumptions:
  + The system will be internally based within one organisation, which uses the hardware and software required by each sub-system and module.
  + Each sub-system and module will have fast and secure communication methods will all other sub-systems and modules.

### System Structure

The service’s structure can be broken down into 6 main sub-systems:

#### The Database

This is where user, sheet, feed and item information is stored. It is used by both the front-end and back-end components and acts as a repository between the two.

#### The Front-End

This is the interface of the system. It consists of a number of web pages accessible by users and administrators. It provides the means by which the users and administrators interact with the system (including the back-end). This interface provides methods for users to manage and view feeds, as well as managing the layout used to display these feeds. The interface also provides methods for administrators to manage users, feeds and items, as well as allowing them to manage and monitor the back-end processes.

#### **The Back-End**

This is an application that is continuously running in the background. It provides the main processing of the system. Its main role is to crawl and parse feeds and content. To perform this role this component contains a continuous queue holding references to all feeds and ordered according to crawl priority. The parsers mainly operate on RSS and HTML items, but other content sources (such as weather feeds, forums and social networks) can also be parsed for storable content.

Additional non-essential components of the service are the API web service and the downloader application. The downloader application (and other possible applications) accesses the system by using the API web service as a gateway.

#### The Downloader Application

…

#### The Web Service

…

#### The Shared Database Library

The Shared Database Library is a C++ library used by the frontend, backend and web service sub-systems as a standard interface to access the Newsfeeder database. The library itself provides methods that can be called to access and modify the database entities and records via SQL statements. This library uses the MYSQL++ library to perform standard database tasks (such as executing a query) and instead focuses on generating application specific queries and retrieving the results.

This system was added to ensure all database calls within the system are performed in a standardised way and as such will achieve the same results. It was also added to separate the database connection/query generation from the other sub-systems.

As the library needs to be used by the different systems which are written in different programming languages (PHP for frontend, C++ for backend, Java for the web service), the library provides a number of wrappers/interfaces for each specific version (these are generated by SWIG, Simplified Wrapper and Interface Generator).

## C:\Users\aron\Desktop\desktop\Final Project\newsfeeder\trunk\Documentation\Diagrams\Requirements News Feeder Overview Diagram.pngUsers

There are 5 users of the News Feeder service:

### Standard User

The main end-users of the system, including anyone in the general public. They have registered with the service and can specify feeds they wish to view. They can set a number of user preferences and settings regarding what content they wish to display and how they wish to display it (such as content filters and sheet layouts). They are expected to have basic computer knowledge (sufficient for using a web browser).

### Subscribed User

Standard users who have subscribed to the system for a fee. As such they gain additional benefits related to their use of the system, such as access to custom sheet layouts and access to the downloader application.

### Administrator

Managers of the service. They manage a number of system features and components (such as user accounts and items) and are able to manage and monitor system processes (such as starting and stopping the crawlers). They are considered to have more in depth computer knowledge than a standard user and understand the internal workings of the system.

### Guest

Anonymous users that can register to the service in order to become standard users. Guests can access the main page, FAQ and are also able to view links sent by registered users of the system (such as a link to an article).

### Web Master

Owner of an external site that is being crawled. They access the system via a portal referenced in the crawler HTTP headers in order to specify how frequently they prefer their site to be crawled (or if they want it to be crawled at all).

### **User Environment**

All users of the service will have access to an active ADSL internet connection in order to use the service, with the exception of the Downloader Application which will only require intermittent access to retrieve content.

The users will access the service using one of the following operating systems: Windows, Linux or Mac OSX. For online service access, users will be able to use: Internet Explorer 7+, Google Chrome, Mozilla Firefox or Safari. Standard web technologies will be used by the system to ensure future browsers will be able to access the system as well

## Scope & Limitations

This development of this service will focus on fulfilling the Base, High and Medium level requirements outlined in the Requirements section of this document.

In terms of limitations, it has a few mainly considering the online accessibility of the service for users and of the content sources.

* The ability for a content source to be crawled and parsed by the service will rely on how it displays its updates (e.g. via an RSS feed). If it does not provide these services or does not provide necessary information (such as a URL link to the article) it will not be able to be stored and used by the system.
* Information provided by the feed will be dependent on the information provided by the actual content source (e.g. we cannot show geo-location information if it is not provided in the article).
* The performance of the service for a user will be dependent on the network connection used to access it (we assume a standard ADSL network connection for the requirements defined in this document).

## Constraints

The News Feeder project and its requirements are limited by the following constraint:

* **Network -** The performance of the service will be dependent on the network connection used to access it. The requirements defined in this document assume a standard ADSL network connection.

## Cost and Pricing

The Newsfeeder system will be available for free to the general public. However, the system will also provide a paid subscription service, where a user will pay a small monthly fee to gain additional features of the service, such as:

* Removal of advertising from the site and sheets
* Access to the Downloader Application
* Increased limit on the number of sheets the user can create
* Ability to create custom sheet layouts

As the system is not being marketed as a product, there will be no additional charges (such as licencing) to potential vendors.

## Quality Ranges

The following defined the quality ranges for the Newsfeeder service in terms of performance, usability and other factors:

* **Availability –** the service will be available 24 hours a day, 7 days a week, with the exception of small amounts of scheduled downtime for maintenance.
* **Usability –** the service will be easy to use for both the standard users, who are expected to have only small amounts of technical knowledge, and for the administrators, who are expected to have general knowledge of the system and its components.
* **Maintenance –** the system is able to be changed/configured without the need to reconstruct and redeploy the system.
* **Accessibility –** the service will provide help for the users, both in form of online help (and FAQ) and user and administrator manuals.

For additional quality ranges in terms of performance please see the Non-Functional requirements section of the Requirements document.

## Precedence and Priority

The functionality discussed in this vision document should be available in the first release of the service to the public. Any additional features of lower priority will be added after the deployment over time (provided they can be integrated with the existing system causing minimal disturbances).

For details on the priority of the functionalities and the order they are to be implemented in is discussed in the Priorities section of the Requirements document.

## Competitors and Alternatives

Aside from the alternative of manually retrieving online content from multiple sources, the system has to compete with a number of pre-existing online RSS aggregators. The following are the three most popular competitors, they all provide the same basic functionality, retrieving online content and compiling it in an easy to access location. These competitors and their main differences are:

#### Google Reader

This solution’s main advantage is that it employs an interface common to other Google services (e.g. Gmail) with simple, easily understandable and powerful features. It provides a folders option to read and group subscriptions easily and also provides a simple method for sharing feed items, by email or directly via Google Reader itself. It can handle almost any type of feed and retrieves updated content quickly.

#### Bloglines

This solution differentiates itself from the others by providing the ability to support to publish any feed items (such as clippings or posts) to a personal blog. It allows for blogs to be created directly from this service, allowing both feed and blog management in the same interface. Bloglines does not provide the social media integration and tagging available the other competitors. Additionally this solution provides no OPML export functionality and updates its content the slowest out of these solutions.

#### Newsgator

This solution provides both a desktop client and web version, which sync together to retrieve updates and changes. It allows easy migrations to the service by allowing feed subscriptions to be imported using OPML. The disadvantages are that it does not allow feed reordering and the desktop client is Windows only.

Overall, these competitors all perform similar functions with some advantages and disadvantages. Newsfeeder is designed to implement as many of the successful features of these competitors as possible. The main area in which the Newsfeeder service differs from these competitors is the way the content is displayed. These competitors display the feeds in standard list/inbox type formats, whereas Newsfeeder will display them in fully customisable layouts, reminiscent of standard news sheets. Users will be able to design the layout format, as well as the formatting for the content (such as, are images displayed, how many items from that feed should be displayed).

## Documentation Requirements

The following documents are required for the full system release:

* **Technical Documentation –** This describes the structure of the system and its development process, to enable future developers to make changes to the service.
* **Standard User Manual –** This is provided as a soft-copy from the service website and provides step-by-step processes for how a user can use the service (such as feed management, sheet management, registration, etc).
* **Administrator User Manual –** This provides step-by-step processes in regards to performing actions specific to administrators in the system (such as managing users and starting/stopping modules). It will be provided as both a hard and soft copy to the administrators of the system.
* **Online Help –** This is provided both explicitly, via a set of FAQ’s accessible directly through the online service, and implicitly, via tool-tips, standard icons and a standard interface.