OPAALS PROJECT

Contract n° FP6-034824

WP9: Communication and Dissemination

D9.5.1 - 1st OPAALS Newsletter

Project funded by the European Community under the “Information Society Technology” Programme

Contract Number : FP6-034824

Project Acronym : OPAALS

Deliverable N°: D9.5.1

Due date: Month 12, June 2007

Delivery Date: Month 12, June 2007

Short Description: The system used to generate Opaals Newsletter, thanks to data retrieved from the OKS

Author: Thibaud Desodt

Partners contributed: Techideas, LSE, T6

Made available to: OPAALS Consortium and European Community

Versioning		
Version	**Date**	**Name, organization**
1	10/05/2007	Thibaud Desodt - TechIDEAS

Quality check

Internal Reviewers: Neil Rathbone, LSE

Dependences:

Work Packages	Indicate which WP could benefit from this Deliverables and if possible with planned tasks, including a clear justification
Partners	indicate which partner could benefit from the reading of this deliverable
Domains	news feeds, communication, publication ...
Targets	specify which is the target of this deliverable (Domain researchers, public administrations, SMEs, Scientific communities on specific domains, etc.)

The content of this article is also available on OPAALS Wiki (http://wiki.opaals.org/WP9-D9.5.1)

D9.5.1 - 1st OPAALS Newsletter

Author : Thibaud Desodt (TechIDEAS)

The **Opaals Newsletter** is part of the 9th Workpackage ("Communication and Dissemination") of the Opaals project. As most of the tasks and deliverables in this workpackage, its main goal is to help spreading the fame of the Opaals, as well as other European projects. In this article we will present the Opaals newsletter communication tool that we have developed, and to explain the choices we did at the design-time.

Contents

1. Presentation
    1. What is the Opaals Newsletter?
    2. Functionalities
    3. Qualities
2. Implementation : a news feeds aggregator
    1. Based on web feeds
    2. Architecture
        1. Data sources
        2. Feed Aggregator - output generator
        3. E-mail sender
    3. Data-processing
3. Conclusion
    1. Interests of the solution
    2. Current State
    3. What's next?
4. More about Opaals Newsletter
    1. Configuration
    2. Templates
    3. Filters
    4. Sources
5. See also

Presentation

What is the Opaals Newsletter?

The newsletter is a functionality that is part of the dissemination process of the OKS. It should allow people (within the OPAALS project, but also external visitors) to stay informed about the OKS, by sending on a regular basis the latest news circulating among the OKS about interesting topics. Contrarily to the OKS, the Opaals Web-office should provide a **passive channel of information**. Users would basically receive news, instead of getting connected by means of the Opaals OKS tools. The idea is rather simple: a member of the community, or just someone interested in Opaals' advancements, may wish to stay informed. He could then register to a newsletter and, once a week, or once a month will receive the news regarding the topics he is interested in.

Functionalities

What should it be able to do?

We may draw up a list of needed functionalities in order to fully understand and describe what we expect from the newsletter.

First of all, we should be able to generate a newsletter with informations coming form several data sources, such as the ones already provided by the OKS tools, but also from external sources (information from external collaborators).

A visitor should be able to subscribe to the newsletter in an easy way. He may just have to type in his e-mail address, and topics he wants to be notified about, as well as the delivery frequency.

Qualities

What do we expect from it?

The OKS already accounts for a huge amount of data, and several tools which use is more and more democratized. Therefore, we do not need any new tool to write the information to be broadcasted. The Opaals-newsletter should be generated from data sources of the OKS (wiki, blogs, forum, etc.). Therefore, as the OKS evolves it should be **extensible**, in order to add some new data sources. To be really useful, it should be able to take as an input data in a given standard format.

One of the major needs is **configurability**. The newsletter should allow for total customization, so that it smoothly integrates with Opaals' identity. First of all, we should be able to send e-mails in **various formats**, such as HTML, plain-text, or in any given format. It could then be easily adapted to the Opaals' Branding conventions. Another configurable part is filtering: the system should be able to retrieve only the news in relation with the user's interest, which means the content should be easily filtered. To deal with the necessity of evolution, new components could be plugged in: some new features, such as new templates for output format, or filters, may be easily added. The email-sending should be **safe and efficient**, with a perfect control of the broadcasting process.

Implementation : a news feeds aggregator

This section deals with implementation details. System administrators and developers may be interested in it, but if you are just a bit curious, you may read it too.

Based on web feeds

Our solution is based on **web-feed retrieving**. For this reason, before speaking about the architecture of the newsletter system, we will present this technology.

Web feeds are a data format designed for frequently updated websites that would like to inform users about **last changes**. It is particularly used for the publication of news about a given resource.

In concrete, a web feed, is a structured file containing a list of changes usually accompanied by a date of publication, a description, information about the author as well as a link to see the full resource. Some specific implementations may provide other information.

Although web-feeds are now always **XML-based** (XML is a structured data format), there is currently no standard format for web-feeds. The most famous and used ones are **RSS** and **Atom**, but web-sites implement several versions of these formats.

- *examples* : RSS 2.0 feed of the BBC - Atom feed of Ubuntu

These formats are very used, and a certain number of tools have been developed to make use of it. Therefore, they are being progressively integrated in web-browsers, such as Firefox, or e-mail clients, such as Thunderbird. The most common tools are called Feed Aggregators, and allow to subscribe to a feed, downloading automatically all the news.

As they are based on XML, they are all user-readable, but they can be also easily processed by computers to extract and transform data, which is exactly what the newsletter generation is about. It is a good choice as most of the tools of the OKS already generate news feeds. It also integrates perfectly with the ideas and developments of the Opaals Web-Office.

- *examples from the OKS* : last changes in the wiki RSS 1.0 feed, Atom feed - last posts on Mehita's blog : RSS 2.0 feed

Architecture

What are the different components?

The implemented Newsletter generator counts with several elements :

- one or several **data sources**
- a **feed aggregator**
- an **output generator**
- an **e-mail sender**

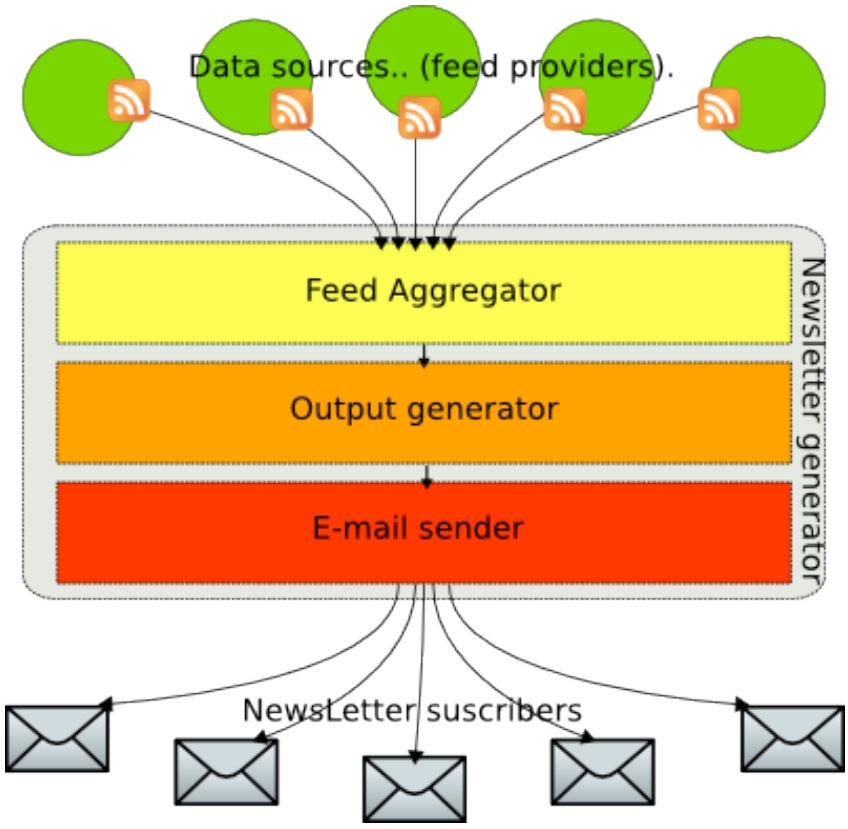

Illustration 1: Overview of the newsletter generation process

The following paragraphs give deeper details about each of these parts.

Data sources

As stated before, data sources must **provide information as news feeds**. This feed generation can be a native functionality, or the website source could be adapted to generate such feeds. The point in such a design is that it allows to add easily new datasources.

The NewsLetter and the Opaals Web-Office were designed to work efficiently in collaboration. It will allow to use a unique sink (the web-office aggregator) to obtain all the data concerning the OKS.

Feed Aggregator - output generator

We decided to use an existing **open-source solution for the feed aggregation and the generation of output** in a convenient and flexible manner : it is called **Planet** (the version we use is actually Planet Venus). It contains both the feed retrieving part and the output generation, and offers many interests.

First, it allows to retrieve data in **nearly all feeds formats**, which means it can retrieve news from nearly every website, and transform it in an homogeneous format.

Content-filters can be plugged-in in order to have a better control of the data we publish. For instance, the published feeds may be filtered so that only the news containing a given word would be sent. We may also apply category filtering, so that (for instance), only the changes in wiki pages of a given category would be broadcasted to the newsletter subscribers.

It provides powerful **outputting possibilities** : the retrieved data can be aggregated and published in any format.

For development concerns, this choice also matches our expectations. It is coded in **Python**, and therefore integrates well with all the development environment of the OKS. It is open-source, which means it can **easily be adapted** to our needs.

E-mail sender

To send emails, we use the Linux **Sendmail**, which is known to be very efficient and very reliable.

Once covered the global architecture of our system ,the next paragraphs explains the process of newsletter generation.

Data-processing

The successive steps in the execution of the newsletter generator are the following:

1. **Read feeds list** from configuration files
2. **Retrieve** news feeds from data sources
3. Convert feeds in a **universal homogeneous format**, trying to guess missing information
4. **Apply filters** to extract the interesting entries
    - category
5. source
6. date
7. author
8. Apply **template and/ or transformation** to generate output files
    - html
9. plain text

10. **Send** the outputted message to the recipients list

Conclusion

Interests of the solution

With such a tool, Opaals is getting ready for communication and dissemination of its community-generated data. It allows to use all the existing infrastructure to provide the public audience with the last events of the Open Knowledge Space.

One of the major interests is the **filtering possibility**. *We can decide to generate a newsletter with only* **changes made by a given user** *(the newsletter administrator for instance), and/or* **on a given topic** *(using category tags)*

Current State

For your eyes' pleasure, here is an example of the output that can be achieved by Planet.

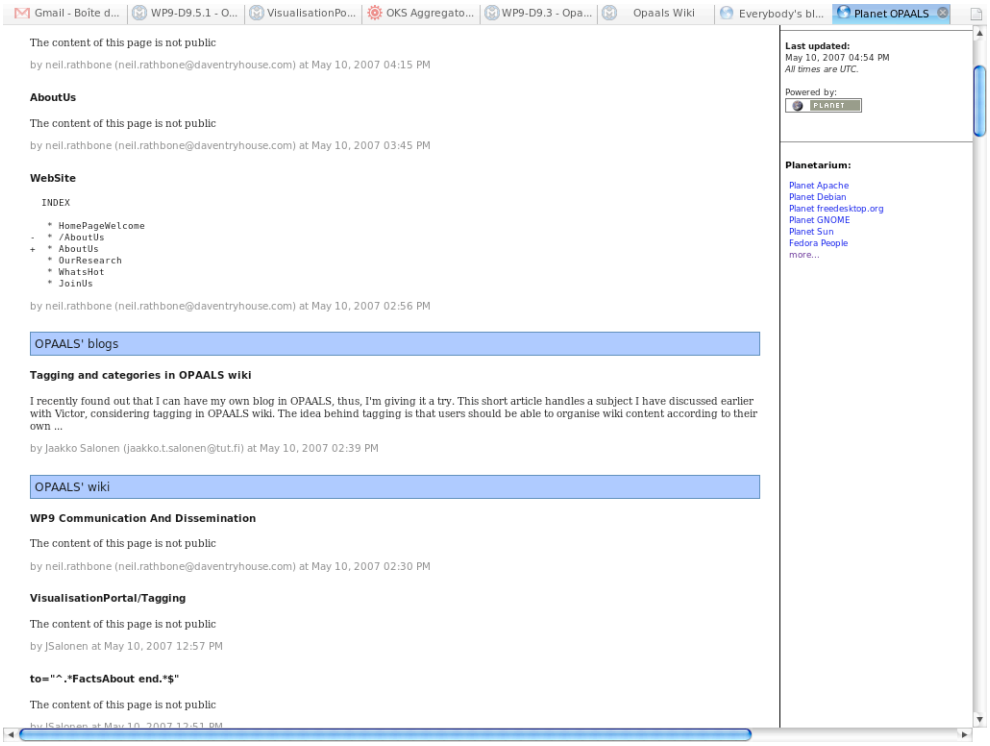

Illustration 2: Screenshot of an obtainable newsletter output

What's next?

- Definition of convention for data structuration
- Applying Opaals Branding
- Make people know Opaals
- Integrate the subscription to Opaals Web-Office

More about Opaals Newsletter

This section contains some more information about the newsletter generator. These are mainly

technical details about configuration and use.

Configuration

Most of the configuration is included is inside a simple plain-text .ini file. This is where you choose :

- the feeds to retrieve
- the templates you want to use for output
- the filters you want to apply

Templates

It is easy to generate output in any text-based format. The easiest way is to use HTML templates, where the program will simply dynamically replace some parts to include the news elements. For XML generation, the best way is to convert from the feed data type thanks to XSL Transformation Stylesheet.

Filters

Filters can be applied on the data we extract from the sources. It allows to customise the information to disseminate. Home-made filters can be implemented, and can allow to select feeds according to the category they belong to, their date of publication or their author author.

Sources

Planet can retrieve nearly any web feed format. It then converts it to a generic feed format which will be used for filtering and processing.

See also

- Opaals Web-Office
- Web feeds on Wikipedia
- XML on Wikipedia
- RSS on Wikipedia
- Atom on Wikipedia
- the Planet project and its Venus branch