



Business Solution

Monitoring and Collection of Information

Product

AMI Enterprise Intelligence

Context

The National League Against Cancer (L.N.C.C.), represents 101 Departmental Committees. It has three principal missions:

- Action for Patients
- Prevention and Information
- Research

The reliable collection of accurate and relevant information is the pivotal activity of the Resource Centre making it a definitive source of verified information able to provide accurate answers to questions right across a highly complex subject .

The L.N.C.C resources centre was directed to develop a policy and process for the collection, analysis and validation of information concerning not only medical research on cancer, but also prevention and assistance for patients

Automation of the Intelligence gathering process

The objectives of the source monitoring and data collection were as follows:

- To be quickly and reliably informed on all aspects of the wide ranging problems associated with Cancer.
- To anticipate requests from members of the public relating to the subject of Cancer
- To update the L.N.C.C resource centre database.

The automation of the data collection process aims to satisfy three requirements:

- To automate the monitoring of information sources across related fields.
- To allow more time for human analysis of the information collected.
- To unify the data gathering process with the subsequent management and distribution of the information.

Detailed attention was paid to the choice of the data sources to watch and the methods of monitoring that were applied to each one of them. The automation of the data collection process takes a different approach to the current manual process.

The following table compares the two approaches:

Manual Monitoring	Automated Monitoring
Newsletters and Distribution lists E-mail alerts for Periodical summaries Examination of Periodical papers Daily consultation of internet sites	Monitoring of News and Press releases Front page or title of a Periodical Determine whether the entire contents of a site should be monitored

Implementation of the solution

The solution used for the automation of data collection, AMI Collect, part of AMI Enterprise Intelligence organises the collection plan in the form of a hierarchy on two levels: Subjects and Scenarios.

The five principal subjects selected were as follows:

- Prevention and Tracking
- Public Health
- Treatments
- Cancer Professionals
- Legal and Social

The drafting of the scenarios follows preset rules, in particular:

For more general sites, for example the Ministry for Health and Family the scenarios were precisely defined to target the most relevant information.

For specialised sites, for example the Curie Institute and the National Cancer Institute, the objective is to be alerted to all new innovations.

Particular care was taken in drafting and creating the scenarios using terminology specific to the field and in certain cases by adjusting the scenarios, such as, for example, Health insurance, cancer = long term effect.

Moreover, the construction of the monitoring plan was carried out in an incremental way: At the beginning, the scenarios were set voluntarily broad in order to guarantee comprehensive coverage; Over a period of four weeks monitoring the results of the collected data these scenarios were refined either to be more precise in description or by varying the relevance thresholds of respective scenarios.

In order to maximise the reliability of collected information and to ensure the quality of research, detailed attention was paid to the qualification of the sources to which the data collection operations were applied.

A particular classification of the sources was made by distinguishing the “primary sources”, for example the website of an important research organisation or a scientific “peer-reviewed” periodical from the “secondary sources” for example a medical news service. After more than one year in daily use with no service interruption the addition of three further processes is being considered:

- The creation of more scenarios relating to further defined research topics ;
- The identification of sources which do not deliver useful information with a level of proof such that these can be confidently omitted.
- To extend the automated data collection to other types of sources, in particular general public sites and newsgroups and compare the results.