



Software Product Description

AMI Market Intelligence v3.0

SPD-AMIMI-30 v1.0

January 2005

Table of Content

1	Approval	3
2	Principles of Operation	3
3	Functional description.....	5
3.1	Organizing: the intelligence map	5
3.1.1	Information sources	7
3.1.2	Conditions (query expression)	7
3.1.3	Collection scheduling.....	8
3.2	Information collection.....	8
3.2.1	Managing information sources.....	9
3.2.2	Setting up connectors.....	10
3.2.3	Query analysis and hypothesis generation	12
3.2.4	Document scoring	12
3.3	Analysis of collected information	13
3.3.1	Statistics	13
3.4	Publishing relevant information	14
3.4.1	Notification.....	14
3.4.2	Instant querying	15
3.4.3	Navigation.....	15
3.5	Advanced linguistic functions	15
3.6	Other functions.....	16
4	Technical environment	17
4.1	Kernel	17
4.2	Specific procedures	17
4.2.1	Man-machine interface.....	17
4.3	Connectors and descriptors.....	18
4.4	Supported environment.....	18
5	Documentation	18
6	Contacts et support.....	19

Please contact us to obtain the latest update of this document.

Albert Meaning Interpreter™, and “What you mean is what you get” are registered trademarks of Go Albert France.

Sherlock is a registered trademark of Apple Computer Inc.
All the other product names are registered trademarks of their respective owners.

Preamble

This document constitutes the functional description of the product:

AMI Market Intelligence v3.0 version 3.0 (reference SPD-AMIMI-30).

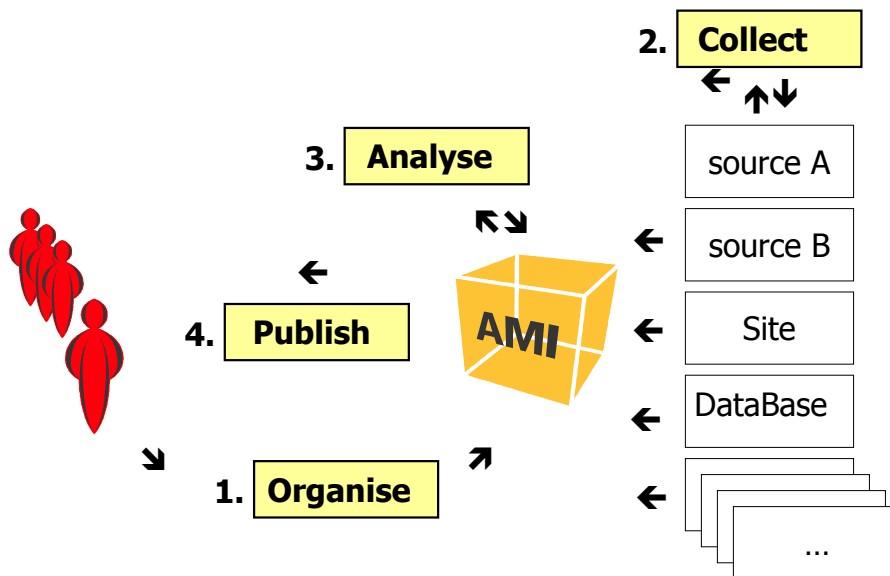
1 Approval

Approved by	Function	Date
Alain Beauvieux	Director of Sales & marketing	January, 6th 2005
Eric Fourboul	Director Products & developpement	January, 6th 2005

2 Principles of Operation

AMI Market Intelligence v3.0 is a software product dedicated to support and automate a market, competitive or technical intelligence process¹.

The Intelligence process



¹ For the purpose of, for example, market, competitive or technical intelligence

The software solution provides support for each of the four major phases of such an intelligence process:

- ✓ The **organisation** of the process;
- ✓ the **collection** of information;
- ✓ the **analysis** of the collected information;
- ✓ the **publishing** of information.

Chapter 3 below describes this process further.

Components of the software solution

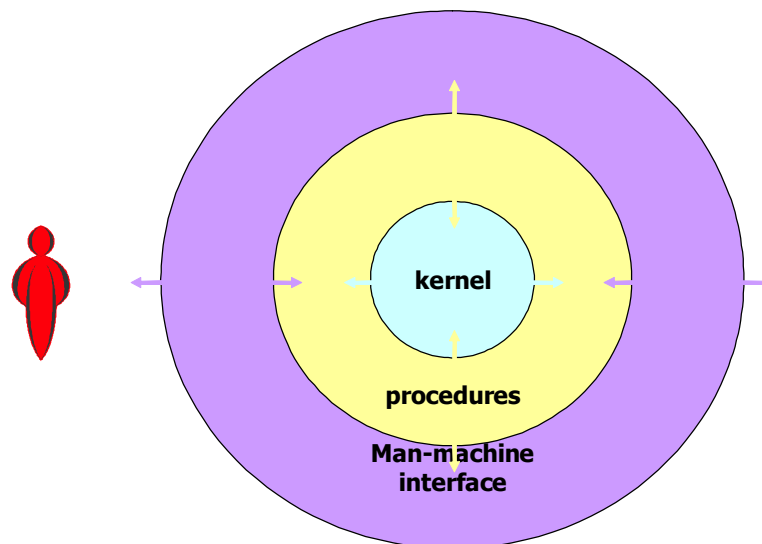
The solution is composed of a kernel which is common to all AMI products, of processes dedicated to the intelligence activity and of a man-machine interface.

The most advanced Go Albert technologies are implemented within this kernel: linguistic processing, computation of a signature, document indexing, etc.

The processes which are specific to the intelligence process implement the kernel components via a set of scripts written in AMI's albScript language.

The **man-machine interface** (MMI) of AMI Market Intelligence v3.0 is completely independent of the other two components. It is therefore possible to modify or replace it to integrate the application within the company's information system or to adapt it to the company's graphic chart.

Section 4 details the technical characteristics of these components.



The efficiency of AMI Market Intelligence v3.0 stems from the ability it provides to automatically organise the collection and analysis processes.

Each user can describe in an intuitive way the topics of interest and the information collection scenarios.

The information collected can reside outside the company (domain specific or competitors' web sites, forums, technical newsgroups, news threads, etc.) or within the company's information system (company portal, content management system, newsgroups, mail environments, databases, etc.)

The documents collected are analysed to automatically de-duplicate information, extract salient expressions and identify quotes which are relevant to the query. Other elements can be added such as the author's name, the document score, trends, etc.). They then are appended to the original documents to create the **news** which are sent to the user.

The published information is for the user's personal use only. The collected information remains the property of its owner and are under the responsibility of the respective owners in compliance with their rights of use and the relevant laws in the country.

3 Functional description

AMI Market Intelligence v3.0 is composed of modules dedicated to the management of the intelligence map, the federated search and analysis of documents.

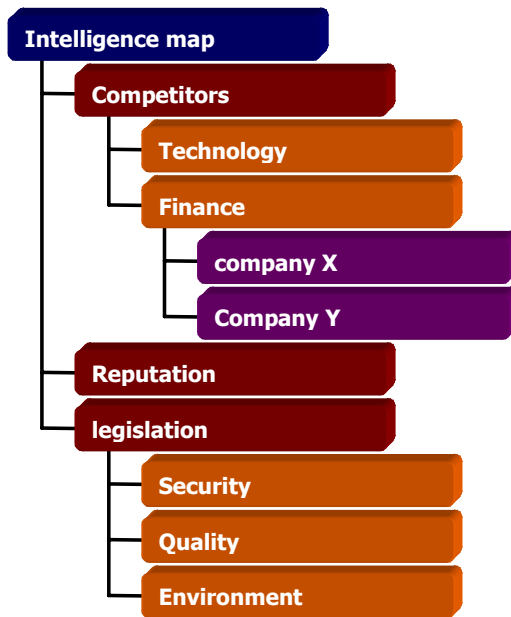
AMI Market Intelligence v3.0 handles the querying process planning and the processing of results, including:

- ✓ **reading** the intelligence map;
- ✓ **managing** the querying of the information sources;
- ✓ **launching the analysis of the documents collected;**
- ✓ **generating an XML structure containing the news resulting from the previous steps which generates an e-mail or is stored in the company's information system.**

A database containing the collected information is build gradually as a result of the intelligence process. This database becomes a valuable information tank containing the history of the intelligence activity thus allowing further analysis and statistics.

3.1 Organizing: the intelligence map

The *Intelligence map* embeds the user's view on the organisation of information. It relies on a tree structure which reflects the company's or user's intelligence activity.



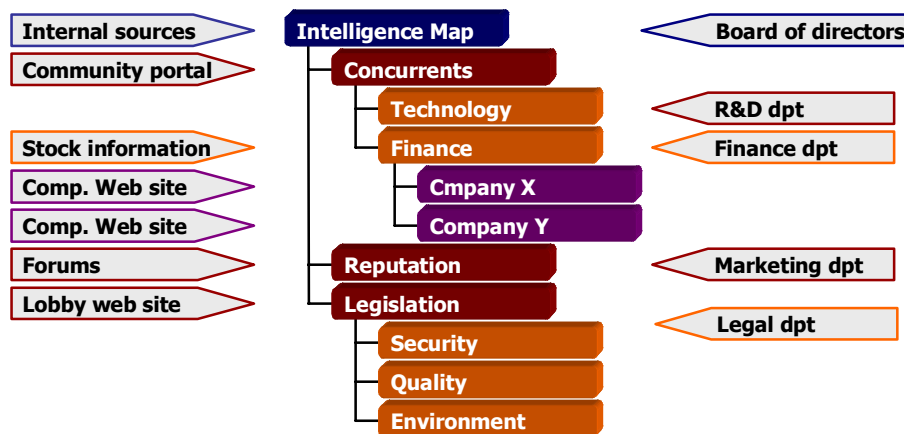
An intelligence map allows not only to describe the way relevant information is collected, but also to categorize information sources.

An intelligence map complies with a dedicated XML DTD (Document Type Definition). The default MMI includes a command to generate such a structure. This structure can also be provided to AMI Market Intelligence by any type of customer specific application.

The following items are included at each level of the tree structure:

- ✓ The source of information;
- ✓ The identifier of the user who will receive the information.

Sublevels inherit from this information. It is therefore possible to cater for any type of organisation. For example:



The intelligence map also allows users to express:

- ✓ The conditions under which documents will be collected;
- ✓ The sub-topic pertaining to a topic;
- ✓ The schedule of information collection and rendering.

3.1.1 Information sources

The information sources which are visible from an intelligence map need to be made available by the system administrator. The system administrator's task consists in gathering several physical sources into logical sources.

There are no constraints on the granularity level of information sources. In the simplest case, there is a one-to-one match between logical and physical information sources. In other cases, physical sources are clustered by themes. For example, physical sources like the MIT, CMU, etc. can be clustered into one logical source name « RESEARCH LABS ».

The management of information sources is described in section 3.2.

3.1.2 Conditions (query expression)

AMI Market Intelligence v3.0 provides users with a powerful and intuitive combination of *strong* and *weak* conditions to retrieve relevant documents.

strong condition: the strong condition is highly restrictive. The terms of the string must be present in the document. This is particularly useful to search information about technical terms or specific company names. A strong condition is useful particularly when the user knows of specific keywords. An alert for the name of a company is an example of a strong condition.

The strong condition can be expressed as a combination of Boolean expression. AMI's Boolean language includes the AND, OR and NOT operators as well as brackets, parenthesis and wild cards. A specific Albert operator (the question mark character) extends the search to all related terms present in the system's knowledge base.

Example: ("Prostat Specific Antigen" OR PSA) NOT (psoriasis OR anorex*)

The query sent to the search engine is calculated by analysing the question submitted by the user. The system generates several search *hypotheses*, each hypothesis corresponding to a query submitted to the search engine or to the information source.

Weak condition: the weak condition is highly permissive. The text that describes the weak condition can be a whole paragraph. The system will search for information similar to this document. For example, a competitor's press release can be used for that purpose. A weak condition is useful when the user wants to collect information which has some resemblance with documents he already knows.

The weak condition can also be used to shallowly define the "universe" in which the information should be looked for. For example the intelligence activity may require to collect information about companies bearing a non-discriminant name (i.e. the company's name is also a common or person's name). In this case, the weak condition can be used to constraint further the search by providing a natural language description of the universe related to this company.

Example:

Stong condition:

Albert OR "Go Albert"

Weak condition:

Information access, online information, competitive intelligence, market intelligence, software, software editor leader...

The weak condition does not require using the Boolean language as the system promotes documents containing the largest number of relevant terms.

3.1.3 Collection scheduling

A field in the structure is dedicated to specify the date range and frequency (daily, weekly, bi-weekly or mothly) for the collection of new documents (see section 4.2, page 17).

3.2 Information collection

The collection of information is an automatic process which is performed daily and controlled by the Intelligence map described above.

The information collection is the most important phase of AMI Market Intelligence v3.0 which will control the whole intelligence process.

3.2.1 Managing information sources

Constraints

The constraints regarding information sources are numerous and varied:

- ✓ Internal/external sources;

Information sources can reside outside the company (like the world wide web) or inside the company. The collection of information can be performed on both areas.
- ✓ Structured or unstructured data basis;

Structured data basis -where information is stored in rows and columns- and unstructured data banks - containing mostly text- collaborate more and more tightly and the border between the two worlds gets thinner. The intelligence system must be able to cater for both type of information sources.
- ✓ Typology of Internet information sources:

Internet is the most common source of information, and is the one which changes the most rapidly. It is also the most accessed source of information amongst users of AMI Market Intelligence v3.0. The Internet is a compound source of information composed of:

 - ⇒ databasis,
 - ⇒ mailing-lists,
 - ⇒ forums, newsgroups, blogs,
 - ⇒ web sites and portals ...

The World Wide Web is therefore not only a huge set of pages in which people navigate, nor is it only a set of search engines. The World Wide Web is an intricate repository of changing information, which can be accessed in numerous ways ranging from a simple URL to a username/password log in, and can require to comply with specific protocols (http, ftp, nntp...). In addition, some sources are explicitly designed to prevent automatic.
- ✓ Complexity and site structure;

Web site navigation, which includes the complexity mentioned above, has been designed for human access and not for surveillance systems. The quality and the relevance of the intelligence process depend on key questions such as: *how can I navigate in the site tree structure? Is it relevant to navigate in directories named "/archives_2003", "/archives_2002"...*? *Should the collection be limited to "news"? Is it relevant to check every day if the document "seminarNov17-2005.ppt" is new?*
- ✓ Freshness of information;

The web site pages which have not changed since the last collection are not relevant for an intelligence activity. This information is though necessary for an automatic system. The http protocol as well as the name of the URL are not reliable enough to filter new documents from old ones. Dedicated algorithms are therefore necessary to perform this filtering without impeding the system's workload.
- ✓ Structure and complexity of contents;

When a relevant page is found, it is important to filter out non-relevant elements such as ads, navigation aids... before analysing the content per se.

- ✓ Availability of service;

The servers on which information reside must be available 24/7. The user is notified when a specific server was not available when the document collection was performed.

- ✓ Web site modification and updates;

As the way a website is organised can change, the collection process must be able to rapidly adapt to a new context.

- ✓ Copyrights issue;

It is important to comply with the copy rights of the document without prohibiting access the collection of copyright free documents.

The response provide by AMI Market Intelligence v3.0 relies on the connector mechanism.

3.2.2 Setting up connectors

Out of the box connectors

Out of the box connectors are provided with AMI Market Intelligence v3.0 to the following search engines:

- ✓ The Yahoo! Search engine, thanks to an international partnership between Go Albert and Yahoo! This partnership gives Go Albert's customers access to:
 - ⇒ Yahoo! International search engine
 - ⇒ Yahoo! News feed which provides daily information from hundreds of information sources
- ✓ The Go Albert engine described in section 4.1.

Indexer and Synchroniser

The *Indexer* and *Synchroniser* are the indexing tools supplying the index tables.

- ✓ The *Indexer* is the **active** indexing mechanism of AMI Market Intelligence v3.0 . It can cover several addresses, carry out an index inversion and update an inverted index. It is active in so far as it proceeds to indexing documents when run.
- ✓ The *Synchroniser*, on the other hand, is the **passive** element of AMI Market Intelligence v3.0. This service is located on an http server. It allows the index to be updated easily by submitting one or several documents for indexing.

Document filters

During the indexing phase, source documents are converted to a common format, in order to submit them to various processes (ex. the discovery of key quotes). At this filtering stage, routines or specific scripts can be connected to guide the analysis, influence the indexing process (ex. to prevent page models from being indexed) or assign to the documents metadata available at the result retrieval stage.

Site crawling

AMI Market Intelligence v3.0 includes functions to crawl external information resources and supports the main protocols in force in this field: http, ftp, file systems, nntp.

Site description

The site description mechanism provides the ability to control the behaviour of the indexer with regards to Web sources. In this way, queries can be addressed to information which are usually not addressed by other systems. These sources include popup pages, java scripts, crawl enhancement.

This mechanism allows also to control how AMI follows links found within a web page and/or indexes target pages.

It also serves to describe which part of a page is relevant in order to extract additional information which will be stored as metadata.

The generic connector

A generic connector is a protocol for querying information resources. It relies on http/XML to ensure communication between AMI and a remote system containing relevant information.

The connection to remote sources is achieved through a dedicated script which ensures a wide range of connection capabilities to sources including the hidden web, dedicated CRM, CMS or publishing solutions.

The generic connector provides the core of AMI Enterprise Discovery federated search capabilities. Virtually any source of information can be accessed, whether they manage structured (such as a database) or unstructured (such as a CMS) information.

Examples are provided in PHP, Perl and Java.

Library of connectors and descriptors

AMI Market Intelligence v3.0 comes with a **library of more than a hundred information sources to the Internet**, including:

- ✓ Official information sources and newspapers;
- ✓ Standardisation offices web sites;
- ✓ Vertical portals;
- ✓ etc.

Reference index

The purpose of an intelligence process is to provide users with information which is new since the last collection. The *reference index* allows to initialize it. Every day (other time intervals can be defined), an optimized automatic process identifies new information thus allowing to make users aware of changes brought to one or several information sources.

This principle is generally combined with the querying process described below:

The system administrator is responsible for managing the collection process and the information sources. Powerful tools are provided as well as a complete documentation.

3.2.3 Query analysis and hypothesis generation

Query analysis

The queries sent to the search engine takes into account both a weak and strong condition as described in scenarios. Several search *hypothesis* are generated, each of which corresponds to a query submitted to a search engine or an information source.

Synonym management

AMI Market Intelligence v3.0 handles synonyms and acronyms. These synonyms may be composed of several words. It is also possible to assign a degree of confidence to the relationship of synonymy. These synonyms are used to generate different hypotheses.

Phonetic and Spelling Approximations

AMI Market Intelligence v3.0 also permits approximate spelling, indeed the words of the user's query are compared to the contents of the knowledge database of the concerned environment. Spelling and phonetic variations are added to the hypotheses generated by the system. This function allows to cater for misspellings in queries and documents alike.

3.2.4 Document scoring

The analysis of the strong and weak conditions allows to precisely define what the user looks for. The content of each document is analysed using powerful algorithms. The proximity between what the user looks for and the document content is reflected in a relevancy score. Documents below a user-defined score are not retained.

Relevancy scoring is described in section 3.5, page 15.

Déjà vu

To ensure that users do not receive the same document twice, a specific function evaluates the newness of documents and discards the ones which have already been delivered to the user.

Duplicate information

AMI Market Intelligence v3.0 evaluates the similarity between the meanings of different documents and can recognize that two documents bear the same information. Depending on the customers choice, duplicated information can be presented or not to the user, or grouped into clusters.

3.3 Analysis of collected information

3.3.1 Statistics

The output of the intelligence process provides a valuable repository containing the history of the information collected according to an intelligence map. Several types of statistic analysis can therefore be performed.

This is made possible by the fact the all the information collected is stored as XML documents in a database which contains the following additional information:

- ✓ The date;
- ✓ The identifier of the leaf in the intelligence map tree structure which allowed to collect the information.

This data can be further analyzed to provide statistic information. Two types of tools are provided by AMI Market Intelligence v3.0:

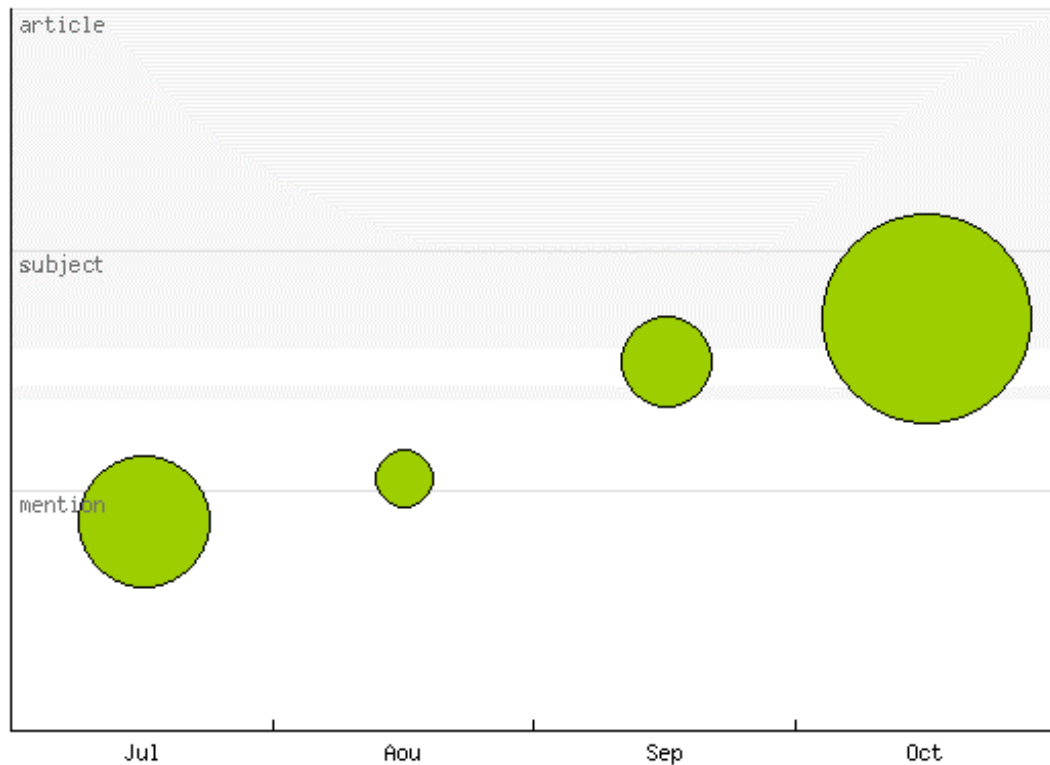
Statistic analysis of the collection database

AMI Market Intelligence v3.0 automatically extracts the main related concepts to identify trends.

Monitoring trends allows to identify in real or deferred time their apparition and evolution on a time scale.

Analysis of the information collected

Dedicated tools allow to analyse the relationship between an intelligence topic and various subjects. The following example shows the relation between the concepts *Iraq* and *The Ivory coast* over a period of 4 months, on the basis of a scenario describing only these two topics:



The system can determine if the collected documents mention the relevant topic, whether the topic is mentioned among others or the document is dedicated to this subject. The volume is represented by the size of the circle. This example (the Ivory Coast scenario) shows that references to Iraq in the Ivory Coast crisis become more and more accurate and that October 2004 has seen a pick., thus confirming the latest news on this issue.

3.4 Publishing relevant information

3.4.1 Notification

The results of the intelligence process are provided in the form of a dedicated XML or HTML structure which can be transformed to fit different types of devices.

The XML structure contains a list of **news**, each news structure contains the following information:

- ✓ URL or address of the main document
- ✓ URL or address of documents conceptually similar to the main document collected

- ✓ URL of the information source (main page)
- ✓ author's name (when available)
- ✓ key-quotes
- ✓ list of relevant keywords
- ✓ identifier of the scenario(s) which retrieved the document
- ✓ meta-data associated with the scenario
- ✓ relevant expressions extracted from the set of returned documents.

This information is usually made available to the user as an e-mail. It can also be made available as an XML structure to be stored in the company's information system.

3.4.2 Instant querying

The information collected over time is stored in a database (see section 3.3.1, page 13), which indexes the documents collected according to the intelligence map, which therefore serves as a categorization structure.

Retrieving documents from this repository is made possible with the **instant querying** function provided by AMI search engine.

All the functions provided by this search engine are available, such as the ones mentioned in section 3.1.2, including weak and strong conditions, as well as everyday and Boolean languages.

Instant querying allows also to instant search within all the sources available to AMI Market Intelligence v3.0.

3.4.3 Navigation

Navigation within the intelligence map gives the user instant access to all the documents collected over time.

3.5 Advanced linguistic functions

The features described below are available in French, German, English, Spanish, Portuguese, Italian and Dutch. These features partly rely on the g-MIL technology (Generator of Metadata Independent of the Language) to compute the **meaning signature** which makes the core of AMI's products' features and significantly improves the relevancy of searches of AMI Market Intelligence v3.0.

Signature

The signature principle allows to:

- ✓ Compute, handle and store a numeric information which represents the actual *meaning* of a text,

- ✓ Support a large variety of signatures ranging from a simple list of terms to a rich and complex set of hierarchized information dependent on the user's context.

Comparing the signatures of two documents allows to measure the distance of the meanings conveyed by the two texts. The more complex the signatures, the more accurate the distance measure. It is still possible to measure the distance between a simple and a complex signature. In the latter case, the system grants a lower confidence to the measure.

Language recognition

Different algorithms are involved in recognising the user's language. These are based on language morphology as well as the user's query history. A recognition confidence score is assigned. Recognising the language in which the query is submitted is necessary to generate adapted search hypotheses.

Key quotes

Key quotes are extracted from documents and automatically stored to help users read results. This is made possible by the system's ability to identify important items within a document. This is performed thanks to the signature described above.

Contextual key quotes

Key quotes containing terms present in the strong or weak conditions are favoured to facilitate the reading of the results returned by the collection process. Note that this function is not available for instant querying.

Trend detection

Trend detection –described in section 3.3.1 – relies on linguistic and statistic algorithms which can be configured by the system administrator.

Stop-word management

A list of stop-words, non-indexed, is provided for each supported language by AMI Market Intelligence v3.0. Furthermore, regardless of the language, during the indexing process of a document by Albert Index, information concerning the importance of terms is gathered to optimise result relevance.

3.6 Other functions

AMI Market Intelligence v3.0 also provides functions to:

- ✓ Manage users profiles;
- ✓ Import/export intelligence maps in CSV format, which facilitates the intelligence maps initialisation.
- ✓ Manage information sources and daily follow up.

4 Technical environment

This chapter provides information regarding the integration of AMI Market Intelligence v3.0 and will be of interest to customers who acquire a product licence. Customers' of the ASP mode can skip this section.

The three layers explained earlier in the section Principles of Operation page 3: kernel, procedures and man-machine interface (MMI).

4.1 Kernel

The kernel is composed of a set of compiled programs which cannot be modified by the user. These programs embed the core of Go Albert's technology:

- ✓ Advanced linguistic functions, including signature computation;
- ✓ Retrieval and relevancy functions;
- ✓ Indexing and set up programs for the connectors and descriptors;
- ✓ albScript language interpreter (as described below).

Most of these programs can be accessed as a web service (http / XML) or from albScript.

4.2 Specific procedures

Specific procedures, which implement all the intelligence functions, are programmed with albScript: AMI's programming language which gives access to whole the functions of the kernel. These procedures provide the basis for a large variety of ad hoc developments. They are used to implement all the intelligence functions, and particularly the **collection** and **diffusion** of information.

They can be technically modified in agreement with Go Albert representative, under certain conditions, in order to extend the system's capabilities:

- ✓ Integrate the collection process to the company's information system;
- ✓ Adapt de-duplication to specific functional needs;
- ✓ Extend the collection frequencies to unsupported values;
- ✓ etc.

The «Programmer's guide» will help users work with albScript. It should be noted that albScript is based on the ECMA-262 (JavaScript) standard. It is open (XML) and provides information exchange with WebServices, external libraries or programs, or ActiveX (COM) modules.

4.2.1 Man-machine interface

The MMI is completely open and can be modified:

- ✓ It can easily be adapted without any constraint to the customer's graphic chart or integrate functions complementing AMI Market Intelligence v3.0;

- ✓ Deeper modifications of the **organisation** or **analysis** functions require prior discussion with Go Albert representatives.

The AMI Market Intelligence v3.0 MMI is available as a set of PHP modules which interface with AMI through normalised XML files.

4.3 Connectors and descriptors

Connectors

Connectors are programs which provide access to information sources and which are developed in any language which complies with the Go Albert html/XML protocol. Tens of connectors which can be enriched and modified by the user are provided as examples.

Connectors can also be developed using the DWU formalism proposed by Go Albert, which is based on Apple's Sherlock protocol.

Descriptors

Site descriptors are defined by the user depending on his needs. Some examples are provided with the product. Descriptors provide all the functions described in section 3.2.2.

A complete documentation on the descriptors and the connectors is provided with the product.

4.4 Supported environment

Please refer to the pre-requisite documentation (réf.: TRFR38140604).

5 Documentation

AMI Market Intelligence v3.0 is delivered with a complete documentation, in French or English, in both pdf and HTML formats, entirely structured. It is composed of:

- ✓ a reference guide
- ✓ a programmers' guide
- ✓ technical prerequisite
- ✓ an administrator's memento

This documentation assists the administrator of AMI Market Intelligence v3.0 in the product installation and everyday administration. It describes all the parameters necessary to control the running of the software, from both technical and functional points of view.

6 Contacts et support

For information relative to AMI Market Intelligence v3.0, please contact the GO Albert company or a local representative referred on Go Albert web site

<http://www.albert.com>.