

# *GathererPLUS*

*Automated Harvesting of Content and Metadata*

Inmagic® *GathererPLUS* is a spidering tool that can assist organizations in addressing their own information glut by crawling specified files servers, Web sites and other content repositories to automatically harvest appropriate content. Today, more and more organizations are focusing on managing their large quantities of internal content - content that can be difficult to manage because it is located throughout their organization, because it exists in many formats, and because it is often dynamic.

*“Inmagic’s software is easy to implement and use. It was up and running in no time.”*

*GathererPLUS*, powered by technology from Amherst Information Group, can be a critical component of an organization’s content management strategy as it allows you to automate the time-consuming process of integrating content from separate silos and repositories, both internal and external to your organization. This content is then loaded into an Inmagic *Content Server* database, indexed based on the meta-data created by the *GathererPLUS* and made accessible for searching and browsing via the corporate intranet. The customer determines the specific files to be crawled and Inmagic *GathererPLUS* crawls the sites, extracting content, and adds metadata tags and other information to further categorize the information.

## **Automated Metadata Extraction**

The *GathererPLUS* includes the following components:

- ◆ Content extraction and normalization engine
- ◆ Metadata extraction for entities such as company names, people, place names, and even customer- specified names or phrases
- ◆ Optional content gathering plug-ins for content in other corporate repositories such as e-mail, FTP or Notes databases
- ◆ Automated and trainable classification [coming soon].

In addition to the standard content formats found on corporate file servers and Web sites - HTML, text files, MS Office documents, presentations and PDF files - *GathererPLUS* offers optional plug-ins that can crawl other content repositories typically found in many organizations such as e-mail servers, ODBC or FTP files and even vendor-specific content such as that stored in a Lotus Notes or Documentum database.

## **Entity Extraction Based on User Defined Lists**

*GathererPLUS* also performs a number of entity extractions to allow creation of additional valuable metadata. Using Inmagic *GathererPLUS*, customers can quickly and easily create searchable databases from the content that is often physically accessible but logically lost within their organizations.

These user defined lists of entities to extract could include a list of product names, or - in a healthcare setting, a list of diseases or - in a financial setting, a list of fund names. You create the list and *GathererPLUS* extracts the information from the content it crawls.

### **Integrated with Inmagic's *Content Server* platform**

When you use Inmagic *GathererPLUS*, you have a set of technology tools that allows you to harvest content from internal data repositories and Web sites and automatically categorize this content based on the topic areas covered by your own organization's data. Since the *GathererPLUS* is integrated with the Inmagic *Content Server* platform, this content is then automatically loaded into your database allowing users to access it with Inmagic *Content Server's* rich set of indexing and searching capabilities. By including user-defined lists, you allow end users to search the database based on the terms included on the list. This can provide valuable access to otherwise hard-to-find content resources.

Whether you are an information professional, an IT Manager or simply one of many involved in developing content management solutions for your organization, Inmagic *GathererPLUS* is a valuable tool that can provide quick and immediate benefits. By allowing you to quickly and easily gather the content resources you need to manage, and integrating internal and external content, *GathererPLUS* can provide you an immediate return on you investment. Combine this with the automated extraction of valuable meta-data, providing you with further categorization of the content, and you have a product that offers clear benefits in many settings.

### **System Requirements for the *Gatherer***

#### Hardware:

- ◆ Intel Pentium processor, 90 MHz minimum
- ◆ 384 megabytes RAM minimum (512 RAM recommended)
- ◆ 240 megabytes hard drive space for application
- ◆ Additional hard drive space for gathered content

#### Operating System:

- ◆ Windows 2000 Professional with Service Pack 3 or Windows 2000 Server with Service Pack 3 or Windows XP Professional

#### Web Server:

- ◆ Microsoft Internet Information Services (IIS) 5.0 or later with Active Server Pages (Only required if you plan to display information using *Content Server for Web*)
- ◆ Microsoft Internet Explorer version 6.0 or later installed on the server where you install the *GathererPLUS Importer* (to support XML functionality)

#### Other Software

- ◆ Inmagic *Content Server* v1.1 or later
- ◆ Adobe PDF IFilter for extraction of PDF documents

## **INMAGIC®**

Inmagic, Inc.

200 Unicorn Park Drive  
Fourth Floor  
Woburn, MA 01801

Tel: 800-229-8398 | Fax: +1 781-938-4446  
E-mail: [marketing@inmagic.com](mailto:marketing@inmagic.com)  
[www.inmagic.com](http://www.inmagic.com)

