

An Intelligent Information Retrieval System based on Meta-Search Engine (IIRS-MSE)

Software Requirements Specification (Example)

Drumm, Paul M

Change History

Date	Summary	Person
03/29/07	Section 1, Started section2, Added 2 Use cases, started the keywords list	Paul Drumm
3/31/07	Section 3	Paul Drumm
3/31/07	Section 2, add 3.8 ,4 ,5 ,6 , Glossary, added 5 Use cases	Paul Drumm
4/24/07	Updated various sections	Paul Drumm
4/29/07	Changes to various sections	Paul Drumm
4/29/07	Edits made to various sections	Paul Drumm

Table of Contents

1	Introduction.....	1
1.1	Purpose.....	1
1.2	Document Conventions.....	1
1.3	Intended Audience, Reading Suggestions.....	1
1.4	Project Scope	1
1.5	References.....	1
2	Overall Description.....	1
2.1	Product Perspective.....	1
2.2	Product Features.....	2
2.3	User Classes	3
2.4	Operating Environment.....	3
2.5	Design and Implementation Constraints.....	4
2.6	User Documentation	4
2.7	Assumptions and Dependencies	5
2.7.1	Assumptions.....	5
2.7.2	Dependencies	5
3	System Features	5
3.1	Create/View Watch List.....	5
3.2	Create/View Alerts.....	5
3.3	Investigate Markets & Sectors	5
3.4	Search for Financial Data.....	5
3.5	Search internal documents	6
3.6	Search and store analytical articles on the Web.....	6
3.7	Search and store news.....	6
3.8	Search and store research views	6
4	External Interface Requirements.....	6
4.1	User Interfaces	6
4.2	Hardware Interfaces	6
4.3	Software Interfaces	6
4.4	Communications Interfaces	7
5	Quality Attribute Requirements.....	7
5.1	Performance Requirements.....	7
5.2	Safety Requirements	7
5.3	Security Requirements	7
5.4	SOX Requirements	8
5.5	Quality Requirements	8
6	Other Requirements	8
6.1	Legal Requirements	8
	Appendix A: Glossary.....	A1
	Appendix B: Use Cases	B1
	UC 1.1 – Search for content.....	B1
	UC 1.2 – Add document to system.....	B1
	UC 1.3 – Add a user to the system.....	B1

UC 1.4 – Add a new search engine or news feed	B2
UC 1.5 – Create or View Watch List.....	B2
UC 1.6 – Create or Modify Alert.....	B2
UC 1.7 –Create or Modify Research View.....	B3
Appendix C: References	C1

1 Introduction

1.1 Purpose

The SRS document describes the functions and requirements of the Information Retrieval System project in a clear and precise manner. Its purpose is to provide a detailed overview of the requirements and features of the system. Document Conventions

1.2 Document Conventions

No special typographic conventions are used in this report. Acronyms and abbreviations used in the document are explained in Appendix A.

1.3 Intended Audience, Reading Suggestions

The intended audience for this document consists of all stakeholders including designers, architects, developers and testers. All technical descriptions and design can be tracked to the use cases and requirements of the project. The document shall serve as a guide through the initial architecture and design phase as well as development and QA process.

1.4 Project Scope

The proposed system will provide fund managers or investors with all the necessary investment related information needed to help them understand the fast changing markets, make critical investment decisions based on market moves and to manage their funds for better performance.

1.5 References

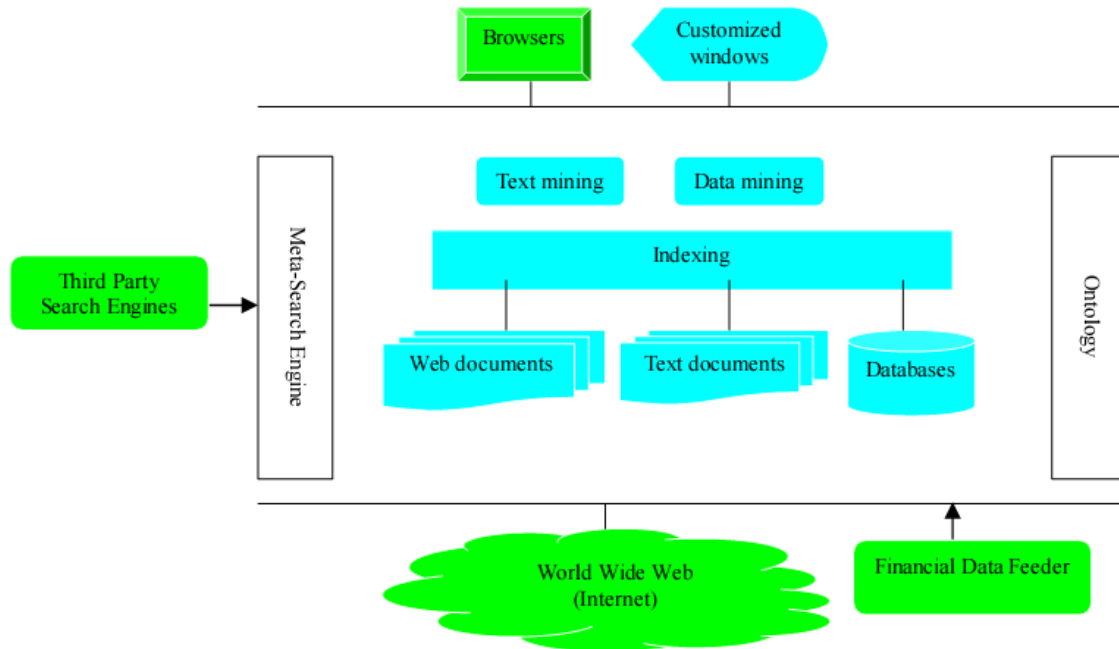
Information Retrieval System Vision and Scope documentation, Vision and Scope.doc

Wieggers, K.E., "Software Requirements, second Edition", 2003, Microsoft Press, ISBN 0-7356-1879-8

2 Overall Description

2.1 Product Perspective

IIRS-MSE is an intelligent information retrieval system based on a meta-search engine and text/data mining techniques. The IIR-MSE also defined as the system, will support fund managers of the investment company and other investors in making investment decisions based on the financial data provided by the system. The system aggregates, indexes, and presents financial data to the end user from a wide variety of sources including third party search engines and financial data feeds. The Ontology, the custom language and reasoning engine interact with the end user queries. The conceptual solution diagram below outlines the major components of the system.



2.2 Product Features

- Financial data feeder will feed the system with historical financial data and news articles.
- Databases will be used to store various investment related data fed by the financial data feeder and the user interface.
- Web documents will be collected by the search engine and by news article feeder.
- Documents will be created and stored by the company in the system.
- The Indexing system will be used for fast text search by keywords entered by the users.
- Ontology will be built and used with the dictionary for better natural language processing for the text documents.
- Meta-search engine, a customized search engine will sends user search queries to several other search engines like Google or Yahoo. It will receive the results from each one, combines, processes and returns the end result to the system.
- The system will provide integration of services provided by the existing search engines in the markets
- Text- mining and data-mining capabilities.
- Browser and stand alone graphical user interface.
- Security and SOX control features.
- Ability to store research meta-data in a hierarchal fashion.
- Ability to send alerts via email after setting thresholds on research entities.
- Ability for Investors to use features.

2.3 User Classes

User Class	Description
Fund Manager	A Fund Manager is an employee of the investment company. This user will be making investment decisions on various funds based on the information presented to her by the system and other sources. A fund manager may also interact with one or several investors as how to interpret the financial data that the investor retrieves from the system and other sources. The fund manager is knowledgeable of the financial business logic associated with various types of funds and the market in general.
Investor	An Investor is a customer of the investment company who has external access to the system. The individual investors have varying degrees of expertise with regards to interpreting the data presented to them by the system and may have to rely on the fund manager for further analysis.
Data Manager	The Data Manager is responsible for managing the system data flows into and out of the system. The data manager is also responsible for managing the functional integrity of the system via various reports about the performance and data flows of the system.
Security Officer	The Security Officer is entrusted with ensuring that data integrity, authentication, and authorization are carried out in a manner as dictated by any regulations and under the auspices of the SOX compliance officer. The Security Officer is responsible for auditing, detecting, analyzing logs, monitoring network traffic and traffic patterns for violations or suspected malicious activities.
SOX Compliance Officer	The SOX Compliance Officer is entrusted with ensuring that all access to and all changes to the system is in accordance with any and all federal, state, and local regulations pertaining to financial data access and presentation.

2.4 Operating Environment

2.4.1 The IIRS-MSE System web servers shall operate on Windows 2003 SP2 running load balanced (WLBS) IIS6 web servers and the .NET 2.0 platforms.

2.4.2 The IIRS-MSE System shall support the following web browsers: Internet Explorer versions 5 and greater, Firefox version 2 and greater, Netscape Navigator version 7 and greater.

2.4.2 The IIRS-MSE System shall support the following operating systems for using the Windowing Form based UI: Windows XP SP2, Vista, and Windows2003 SP2 on the Microsoft Windows Platform. It will also support any Unix based operating system that supports the use of Java Runtime version 6 or greater when in a GUI presentation mode.

2.4.3 The IIRS-MSE System shall support users that geographically reside in USA.

2.4.4 The IIRS-MSE system shall support en-US code pages (English reading).

2.4.5 The IIRS-MSE system shall be located behind the corporate Checkpoint 1 firewall in the DMZ and the firewall must allow HTTP over port 80 to the web servers, HTTPS over 443 to the web servers, and SMTP over port 25 outbound from the web servers. Any users that need access to the Windowing Form based UI will use a Checkpoint Secure-Remote client to attach to the firewall as an extension of the corporate network.

2.4.6 The IIRS-MSE system shall be on a Verizon frame relay circuit capable of sustaining 500 kb/s throughput in both directions at the same time.

2.4.7 The IIRS-MSE system WEB UI shall initially operate on two separate web servers to balance the load.

2.4.8 The IIRS-MSE system database servers will run on two Windows 2003 SP1 servers running a SQL 2005 database cluster in active/passive mode which supports the use of indexing any and all data fed into the system.

2.4.9 The IIRS-MSE system Analysis servers for cube building, data mining, and ontology will be running on two Windows 2003 SP1 servers running a SQL 2005 database cluster in active/passive mode and may be the same as indicated in 2.4.8

2.5 *Design and Implementation Constraints*

2.5.1 The system shall use the current corporate standard Microsoft SQL 2005 Enterprise database engine. The data for analysis in cubes will reside on a Microsoft SQL 2005 Analysis Service to facilitate data and text mining.

2.5.2 All HTML code shall conform to the HTML 4.0 standard.

2.5.3 All scripts shall be written in javascript.

2.5.4 The web site shall be developed in ASP.NET 2.0 using C#.

2.5.5 The machine keys shall be the same on all web servers that control state encryption.

2.5.6 There shall be no ad hoc queries encapsulated in the web pages, all data queries shall be done through stored procedure calls encapsulated in transactions.

2.5.7 The Windowing Based UI will be written in Java to facilitate cross platform data analysis.

2.5.8 All internet based transfers with the browsers have to be encrypted using SSL.

2.5.9 All requests will be audited as to what the request was for, who made the request, timestamp of the request, and if the attempt for the request was successful or if it violated any security policies.

2.5.10 The Ontology and its characterizations should be stored in an SQL database.

2.6 *User Documentation*

2.6.1 The system shall include an online HTML-based help system to describe features, to provide assistance and instructions for properly using system functions. This will be used both for the web browser and the Java based Windowing UI.

2.6.2 The system shall provide an online user manual accessible to users with proper access. The online manual shall include online tutorials.

2.7 Assumptions and Dependencies

2.7.1 Assumptions

2.7.1.1 Investors and Fund Managers will have valid accounts set up by the Security Officer.

2.7.1.2 Fund Managers using the system for standard business operations will be trained to properly use the system features related directly to their roles and responsibilities.

2.7.1.3 Investors using the system for inquiries will be trained to properly use the system features related to what they have the ability to access and before they are given access to the system.

2.7.1.4 Proper levels of authorization and authentication will be established for Investors and Fund Managers to access the system information directly related to their role(s).

2.7.2 Dependencies

2.7.2. The system will depend on the availability of the third party search engines and financial data feeds so no claims can be made as to the accuracy or reliability of the data in a given time frame.

3 System Features

3.1 Create/View Watch List

The user maintains a list of companies that are currently the focus of her research and investigation.

3.2 Create/View Alerts

The user maintains a list of companies and triggers which will result in alerts to be issued to her.

3.3 Investigate Markets & Sectors

The user can categorize and further pinpoint her research and information gathering on particular industries that are part of her funds or are candidates to be added to the funds under her management.

3.4 Search for Financial Data

This feature is used to view current and historical financial data to analyze market position of the company that is the focus of her research.

3.5 Search internal documents

This feature is used to search for information stored in the investment company's internal file storage and database.

3.6 Search and store analytical articles on the Web

This feature is used for a more thorough and deeper investigation of targeted companies using outside sources that would include financial magazines, university research material etc.

3.7 Search and store news

This feature provides the fund manager with the ability to search and store any news published by online newspapers or domestic and international news organizations.

3.8 Search and store research views

This feature provides the fund manager or investor with the ability to group and store various metadata (hyperlinks, feed, document id) about her various searches in a search panel.

4 External Interface Requirements

4.1 *User Interfaces*

- 4.1.1 A CSS style sheet will be retained and used across all web pages.
- 4.1.2 The web pages shall be navigable via keyboard or mouse.
- 4.1.3 Online Help links shall be visible on every page in the upper right corner of the page.
- 4.1.4 The horizontal viewable display width of screen shall be visible in an 800x600 display mode.
- 4.1.5 All pages will use the en-US code page.

4.2 *Hardware Interfaces*

4.2.3 There shall be a high-speed scanner interface to allow importing of physical documents into the system. This shall be kept in a location secured by the Security Officer.

4.3 *Software Interfaces*

- 4.3.1 The system shall be able to interface with multiple external aggregated news feeds using a RSS 2.0 format.
- 4.3.2 The system shall be able to interface to other customized news feeds using the WSDL generated by the vendor feed.

4.3.3 The system shall interface with various third party search engines using the Meta-Search Engine.

4.4 Communications Interfaces

4.4.1 Clients must be able to hold an SSL session with the web servers via HTTPS over port 443.

4.4.2 Clients using the Windowing based UI must be able to open ports to communicate with the checkpoint firewall.

4.4.3 News feeds will communicate via standard HTTP over port 80 inbound using the RSS 2.0 standard.

4.4.4 Notifications via SQL 2005 Notification services shall be sent out via standard SMTP mail over port 25 to end user's email address.

5 Quality Attribute Requirements

5.1 Performance Requirements

5.1.1 The IIRS-MSE System shall not take more than 15 seconds for any page to fully display.

5.1.2 The IIRS-MSE System shall return query results in 15 second or less.

5.1.3 The IIRS-MSE system shall be designed to support 200 concurrent users with an estimated average session duration of 30 minutes.

5.1.4 Any cube builds, re-indexing, or large data imports shall not affect the other performance requirements.

5.2 Safety Requirements

n/a

5.3 Security Requirements

5.3.1 Access to the IIRS-MSE system shall be based on a secured role based system with encrypted passwords between 7 to 14 characters.

5.3.2 The IIRS-MSE system shall encrypt all sessions that are web based.

5.3.3 The security officer shall sign off on any data inputs into the system.

5.3.4 The security officer shall authorize any new news feeds or search engine additions.

5.3.5 All access must be audited and kept in audit history tables for 3 years.

5.4 SOX Requirements

5.4.1 The SOX officer shall ensure that any web page that gives out information to investors be labeled as possible forward-looking statements.

5.4.2 The SOX officer shall ensure that any and all data changes made to the system are physically done by employees who are not directly involved in making the changes to the data.

5.4.3 The SOX officer must sign off on all changes to the system.

5.5 Quality Requirements

5.4.1 The IIRS-MSE system shall be available and running 24 hours a day, seven days a week, and 365 days a year.

5.4.2 The IIRS-MSE system's user interface shall be familiar to users who have used other web applications and Windows desktop applications.

5.4.3 The IIRS-MSE system shall be upgradeable with no downtime or disruption.

5.4.5 The IIRS-MSE system shall be composed of highly structured programs to allow easier separation of parts for independent testing of system components.

5.4.6 The IIRS-MSE system shall provide tutorials and an online/searchable user help function for all user classes.

6 Other Requirements

6.1 Legal Requirements

6.1.1 Any use of the IIRS-MSE system must comply with all federal regulations, directives and guidelines governing the use of investment data.

6.1.2 Terms and conditions of website use must clearly be displayed to all users of the site.

6.1.3 Any data viewed by investors must be labeled as possible forward looking data that may change at any given time.

6.1.4 The SOX or Security officer may make changes to this document at any time to comply with any federal, state, local, and corporate rules, regulations, or other legal directives.

Appendix A: Glossary

Term	Definition
Browser	A web browser is a software application that enables a user to display and interact with text, images, and other information typically located on a web page at a website on the World Wide Web or a local area network. (See Web Browser)
Data-mining	The process of extracting interesting/useful patterns from data (typically in structured format).
Database	Structured collection of records or data that is stored in a computer so that a program can consult it to answer queries. The records retrieved in answer to queries become information that can be used to make decisions. The computer program used to manage and query a database is known as a database management system (DBMS) .
Dictionary	A dictionary is a list of words with their definitions , a list of characters with their glyphs , or a list of words with corresponding words in other languages.
Financial data feeder	Feeds the investment related data, including historical prices, other financial data of companies, and the news articles to the system.
GUI	Graphical User Interface; an output interface used to display images, animation and text.
HTTP and HTTPS	Hyper Text Transport Protocol is a protocol for transferring information across the network over TCP/IP. HTTP also specifies the formatting of such data as it flows. HTTPS is the same but the data is encrypted by virtue of a set of key exchanges that encrypts and decrypts the data. An asymmetric key exchange takes place between the server and client which then includes the symmetric key that is

	used to encrypt and decrypt the remaining data exchange.
Indexing system	Is a system that extracts content from files and other text storage systems and constructs an indexed catalog to facilitate efficient and rapid searching.
IIRS-MSE	Intelligent Information Retrieval System based on Meta-Search Engine.
Keywords	In the indexing process, the gathered text data is passed through a word breaker to separate the text into individual tokens, or keywords which are then matched against a search criteria.
Noise Words	Not all words in a document are worth indexing. To enhance the quality of matches returned in response to a query, you can remove such <i>noise words</i> by creating a <i>stop list</i> . In addition, removing noise words significantly reduces the size of the index, which facilitates faster searching.
Meta-search engine	A customized search engine that sends user search queries to several other search engines like Google or Yahoo, receives the results from each one, combines, processes, and returns processed results to the system.
Natural language processing	Natural language processing enables machines to understand what people mean and respond with intelligence. People benefit from this work because it allows them to interact with a computer <i>naturally</i> , without having to adapt their behavior based on the computer's limitations.
News article	A news article is a meaningful textual representation about a subject or subjects in print or electronic format. When in electronic format it is usually arranged in a hierarchal fashion within a certain subject area called a news group on a news server that uses the NNTP protocol to serve up news articles to the end users news reader.
News article feeder.	Used for serving users frequently updated content. Content distributors <i>syndicate</i> a news feed, thereby allowing users to <i>subscribe</i> to it. Making a collection of news feeds accessible in one spot is known

	as <i>aggregation</i> .
Ontology	A data model that represents a set of concepts within a domain and the relationships between those concepts. It is used to reason about the objects within that domain.
Our	The usage of the term refers to Global Inventory Inc.
QA	Quality Assurance; the process or set of processes or measures used to assure the quality of a product.
Search engine	An information retrieval system designed to help find information stored on a computer system , such as on the World Wide Web , inside a corporate or proprietary network, or in a personal computer .
System	The embodiment of the IIRS-MSE.
Text document	Documents created and stored by the company.
Text- mining	The process of extracting interesting/useful patterns in text documents.
Text search	A technique for searching a computer -stored document or database ; in a text search, the search engine examines all of the words in every stored document as it tries to match search words supplied by the user.
Web Browser	A web browser is a software application that enables a user to display and interact with text, images, and other information typically located on a web page at a website on the World Wide Web or a local area network. (See Browser)
24/7	Referring to operations that are conducted 24 hours a day, 7 days a week.

Appendix B: Use Cases

Use Case ID

UC 1.1 – Search for content

- Goal:** The actor enters a search phrase and the system displays a list of matching results.
- Actor(s):** The fund manager or the investor.
- Preconditions:** The actor has navigated to the search page or opened the standalone search application.
- Trigger:** The actor selects “search”.
- Main Scenario:** Actor enters a search phrase.
Actor selects “Search”.
System displays a list of matching result.
- Alternatives:**
- End Conditions:** User gets a list of matching results.

Use Case ID

UC 1.2 – Add document to system

- Goal:** Actor adds a document to the system.
- Actor(s):** The fund manager or the investor.
- Preconditions:** The actor has navigated to the search page or opened the standalone search application.
- Trigger:** The actor selects “Add document”.
- Main Scenario:** System displays a file dialog.
Actor browses to the file he wants to add.
Actor selects the file.
Actor selects ”Add File”.
System confirms addition.
- Alternatives:** Unknown file format.
- End Conditions:** User gets a list of matching results.

Use Case ID

UC 1.3 – Add a user to the system

- Goal:** Actor adds a user to the system.
- Actor(s):** The Security Officer.
- Preconditions:** The actor has navigated to the ADD USER form in the Window Based UI and is logged in.
- Trigger:** The actor selects “Add User”
- Main Scenario:** System displays a dialog.
Actor enters username.
Actor enters password.
Actor enters base email address.
Actor enters access method either web or windows based UI

Actor enters personal information about user.
System confirms user was added.
Alternatives: Password not complex enough and dialog box explains.
End Conditions: User can log into the system and is alerted when she can.

Use Case ID

UC 1.4 – Add a new search engine or news feed

Goal: Actor adds a new search engine or news feed to the system.
Actor(s): The Security Officer and SOX Officer.
Preconditions: The actor has navigated to the ADD feed/engine form in the Window Based UI and is logged in.
Trigger: The actor selects “Add Feed/Engine”
Main Scenario: System displays a dialog.
Actor selects feed or engine.
Actor enters feed location or engine location using a URI.
System confirms engine or feed was added.
Alternatives: n/a
End Conditions: News Feed or Search Engine Results are available to the system.

Use Case ID

UC 1.5 – Create or View Watch List

Goal: Actor wants to view a watch list.
Actor(s): Fund Manager, Investor.
Preconditions: The actor has navigated to the view watch list page and is logged in.
Trigger: The actor selects “View Watch List”.
Main Scenario: System displays a dialog.
Actor selects Watch List to view.
System displays watch list.
Alternatives: Alternatively actor is given the option to add to the current watch list.
Alternatively actor is given the option to add to a new watch list.
End Conditions: Actor views watch list and/or creates new one.

Use Case ID

UC 1.6 – Create or Modify Alert

Goal: Actor wants to create or modify an alert.
Actor(s): Fund Manager, Investor, Security Officer, Data Manager.
Preconditions: The actor has navigated to the My Alerts page and is logged in.
Trigger: The actor selects “My Alerts”.
Main Scenario: System displays a dialog.
Actor selects alert to modify.
Actor selects objects to be alerted on.
Actor selects threshold at which alert is fired.
Actor selects intervals at which alerts are repeated.

- Actor selects whether to disable alert.
Actor selects which email address on record to add.
- Alternatives:** Alternatively actor is given the option to add a new alert.
Alternatively actor is given the option to delete an alert.
- End Conditions:** Actor creates/modifies/deletes an alert on the “My Alerts” page.

Use Case ID

UC 1.7 –Create or Modify Research View

- Goal:** Actor wants to modify research view.
- Actor(s):** Fund Manager, Investor.
- Preconditions:** The actor is in a search result page, viewing an actual result, or has navigated to the Research View page and is logged in.
- Trigger:** The actor selects “Add to My Research”.
- Main Scenario:** System displays a dialog in a new window listing the research views.
Actor selects Research View to add article id, hyperlink, or news feed to.
System adds it to the Research View selected.
- Alternatives:** Alternatively actor is given the option to make a new Research View.
Alternatively the actor can select ‘cancel’.
Alternatively the actor can select a link in the research view she selected and open that link in a new window.
- End Conditions:** Actor creates, modifies, or views document in Research View.

Appendix C: References

Huzinga & Kolawa, Automation and Defect Prevention in Software Management, Wiley Inc, 2007

Wieggers, K.E., "Software Requirements, second Edition", 2003, Microsoft Press,