

Traceability and Information Retrieval

Arthur T.
Jeton M.
Michael S.
Rita P.
Robert R.

Summary

1. Reminder: Traceability
2. Treaceability and Information Retrieval
3. The Problem
4. Information Retrieval: How it works
5. Automation
6. Quality measures
7. RETRO
8. Conclusion

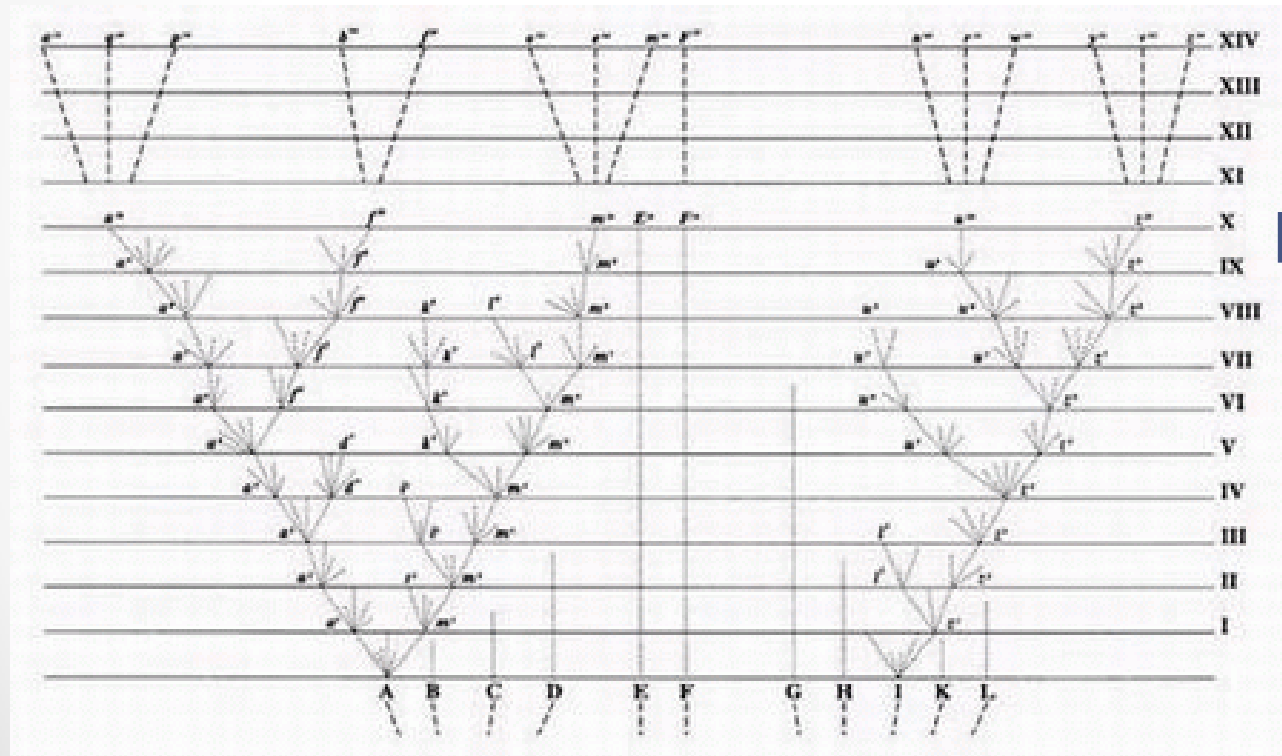
Keep in mind

Requirements Traceability:

Step	Task
(a)	identify each requirement
(b)	assign a unique identifier to each requirement
(c)	for each high level requirement, locate all matching low level requirements
(d)	for each low level requirement, locate a parent requirement in the high level document
(e)	determine if each high level requirement has been completely satisfied
(f)	prepare a report that presents the traceability matrix
(g)	prepare a summary report that expresses the level of traceability of the document pair

The Why of Traceability

- "Know where you come from"



Requirements
Evolution

The Tree of Life, On the Origin of Species by Natural Sélection, 1859

Traceability and Information Retrieval (IR)

The Problem

"But do you know that, although I have kept the diary [on a phonograph] for months past, it never once struck me how I was going to find any particular part of it in case I wanted to look it up?"

Dr Steward, Bram Stoker's *Dracula*, 1897

About Information Retrieval

Standard Problem: Given a document collection and a query, determine those documents from the collection that are relevant to the query.

 Relation to the session's topic!

High-level requirements = the queries

Low-level elements = the document collection

Relation to the session's topic

Information Retrieval (IR) problem is very similar to the Requirements Traceability problem:
Especially in candidate link identification

Information Retrieval: also a matching problem

Information Retrieval

1. User enters a query into the IR-System
2. IR-System matches the query against the DataBase (Knowledge Base)
3. IR-System computes a matching score for each object of the DB
4. IR-System ranks the objects along the matching score
5. The objects are presented to the user in decreasing matching rank
6. The user can refine the query with another cycle

Humans vs. Computers: Automation (1)

Analyst's responsibilities :

ID	Capability
1.	Capturing Requirements/Identification
1.1	Input document enrichment / identification Using existing document information, aid the user in requirements analysis, identification of requirements, etc.
1.1.1	Input document change/comparison analysis The ability to compare/contrast two different versions of a source document
2.	Capturing system element structure
2.1	Graphically capture system structure
2.2	Textual capture of system structure

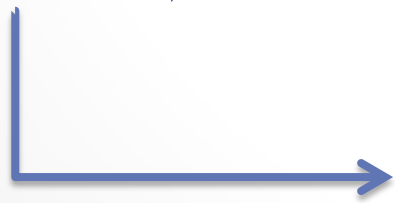
Humans vs. Computers: Automation (2)

PROBLEM: Current approaches are relying on the analyst.

Errare humanum est



Automation



Let the computer do the
stupid, repetitive and
error-prone work



Central point of the
literature

Measures

2 main measures for IR performance:

➤ Precision

➤ Recall

Example: RETRO

The screenshot shows the RETRO software interface. The title bar is labeled "RETRO". The menu bar includes "File", "Action", "Data", "Help", and "About". The toolbar shows "first" and a magnifying glass icon. The main window has two tabs: "TRACE" and "BROWSE".

Annotations with blue ovals highlight the following elements:

- High Level File Names**: A table with columns for file names and a "FREEZE" column.
- Low Level File Names**: A table with columns for file names, weights, and links.
- High Level Text**: A text area containing the following text:

Each MODIS standard data product shall be produced within the data volume and processing load allocation shown in Table B-1.
- Low Level Text**: An empty text area.
- FILTER**: A filter input field with a value of "0.0".

At the bottom, there are radio buttons for "LOCAL" and "GLOBAL", and three buttons: "Trace All", "Trace Current", and "Save".

	High Level File Names	Low Level File Names																																																															
	<table><thead><tr><th></th><th>File Name</th><th>FREEZE</th></tr></thead><tbody><tr><td>0</td><td>SDP3.2-2</td><td>yes</td></tr><tr><td>1</td><td>SDP3.3-1</td><td>no</td></tr><tr><td>2</td><td>SDP3.3-2</td><td>no</td></tr><tr><td>3</td><td>SDP3.3-4</td><td>no</td></tr><tr><td>4</td><td>SDP4.2-1</td><td>no</td></tr><tr><td>5</td><td>SDP4.2-2</td><td>no</td></tr><tr><td>6</td><td>SDP4.2-3</td><td>no</td></tr><tr><td>7</td><td>SDP4.3-1</td><td>no</td></tr></tbody></table>		File Name	FREEZE	0	SDP3.2-2	yes	1	SDP3.3-1	no	2	SDP3.3-2	no	3	SDP3.3-4	no	4	SDP4.2-1	no	5	SDP4.2-2	no	6	SDP4.2-3	no	7	SDP4.3-1	no	<table><thead><tr><th></th><th>File Name</th><th>WEIGHT</th><th>LINK</th></tr></thead><tbody><tr><td>0</td><td>L1APR01-I-1</td><td></td><td>none</td></tr><tr><td>1</td><td>L1APR01-F-4-4</td><td>0.32638</td><td>none</td></tr><tr><td>2</td><td>L1APR03-F-2.4-1</td><td>0.21678</td><td>none</td></tr><tr><td>3</td><td>L1APR01-F-4-3</td><td>0.17473</td><td>none</td></tr><tr><td>4</td><td>L1APR03-I-5</td><td>0.13162</td><td>none</td></tr><tr><td>5</td><td>L1APR01-F-2.3-1</td><td>0.10260</td><td>none</td></tr><tr><td>6</td><td>L1APR03-F-2.4-2</td><td>0.09437</td><td>none</td></tr><tr><td>7</td><td>L1APR01-F-2.4-1</td><td>0.08251</td><td>none</td></tr></tbody></table>		File Name	WEIGHT	LINK	0	L1APR01-I-1		none	1	L1APR01-F-4-4	0.32638	none	2	L1APR03-F-2.4-1	0.21678	none	3	L1APR01-F-4-3	0.17473	none	4	L1APR03-I-5	0.13162	none	5	L1APR01-F-2.3-1	0.10260	none	6	L1APR03-F-2.4-2	0.09437	none	7	L1APR01-F-2.4-1	0.08251	none
	File Name	FREEZE																																																															
0	SDP3.2-2	yes																																																															
1	SDP3.3-1	no																																																															
2	SDP3.3-2	no																																																															
3	SDP3.3-4	no																																																															
4	SDP4.2-1	no																																																															
5	SDP4.2-2	no																																																															
6	SDP4.2-3	no																																																															
7	SDP4.3-1	no																																																															
	File Name	WEIGHT	LINK																																																														
0	L1APR01-I-1		none																																																														
1	L1APR01-F-4-4	0.32638	none																																																														
2	L1APR03-F-2.4-1	0.21678	none																																																														
3	L1APR01-F-4-3	0.17473	none																																																														
4	L1APR03-I-5	0.13162	none																																																														
5	L1APR01-F-2.3-1	0.10260	none																																																														
6	L1APR03-F-2.4-2	0.09437	none																																																														
7	L1APR01-F-2.4-1	0.08251	none																																																														

Conclusion

Effective Requirements Traceability, using an IR tool:

- Semi-automated process
- Machine does the repetitive work
- Analyst makes decisions
- Some other possibilities to further improve: e.g. filters

Questions?