

RE II: Traceability

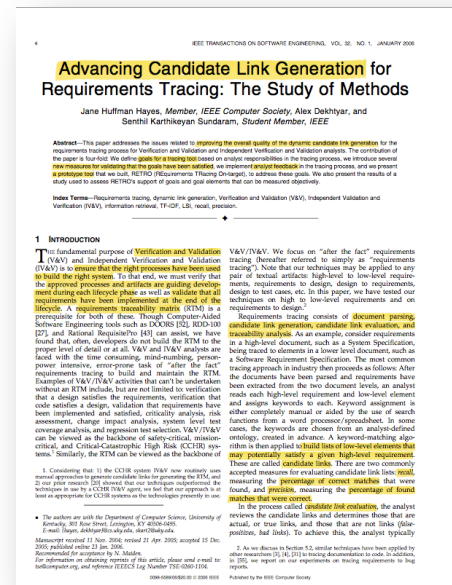
Traceability and Information Retrieval
Christian Anastasiu, Samuel Liniger, Fabian Hofstetter



Literature

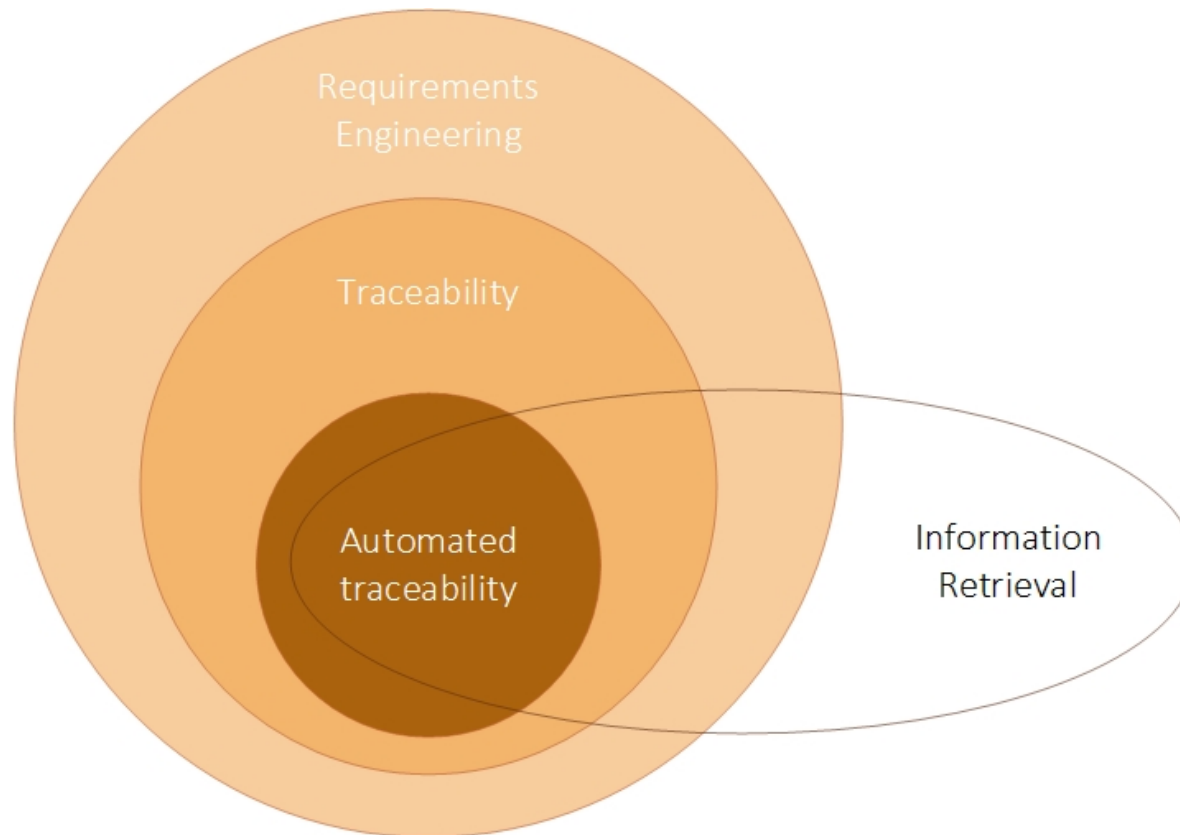


Best Practices for Automated Traceability - Cleland-Huang, Settini, Romanova, Berenbach, Clark (2007)



Advancing Candidate Link Generation for Requirements Tracing: The Study of Methods — Hayes , Dekhtyar, Sundaram (2006)

Context



Definition

- ◆ Requirements Traceability

- ◆ „follow the life of a requirement in both a forward and backward direction“[1]

- ◆ Information Retrieval (IR)

- ◆ “Information retrieval is finding material of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)”[2]

[1] O.C.Z. Gotel and A.C.W. Finkelstein, “An Analysis of the Requirements Traceability Problem,” Proc. 1st IEEE Int’l Conf. Requirements Eng., IEEE CS Press, 1994, pp. 94-101.

[2] Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. Introduction to information retrieval. Vol. 1. Cambridge: Cambridge University Press, 2008.

Requirements Tracing Process



document parsing



candidate link
generation



candidate link
evaluation



traceability
analysis

RETRO

- ◆ Use IR to find „similar“ Requirements for candidate link generation
- ◆ IR methods
 - ◆ Tf-Idf
 - ◆ Tf-Idf + simple Thesaurus
 - ◆ Latent Semantic Indexing
- ◆ IR measures
 - ◆ recall = percentage of true links found
 - ◆ precision = accuracy of the returned candidate link list

RETRO Evaluation

- ◆ Goals

- ◆ Believability (Accuracy, Scalability, Utility)
- ◆ Discernibility
- ◆ Endurability

- ◆ Result

- ◆ Accuracy & Scalability with some methods & some data sets
- ◆ Discernibility with some methods & some data sets
- ◆ Endurability with some methods & some data sets

Best practices

- ◆ Traceability environment
 - ◆ Trace for a purpose
 - ◆ Define a suitable trace granularity
 - ◆ Support in-place traceability

Best practices

- ◆ Creating traceable artefacts
 - ◆ Use a well-defined project glossary
 - ◆ Write quality requirements
 - ◆ Construct a meaningful hierarchy
 - ◆ Bridge the intradomain semantic gap.
 - ◆ Create rich content
- ◆ Introducing automated trace processes
 - ◆ Process improvement plan

Take Home Messages

- ◆ Requirements Traceability can help you do
 - ◆ the right things (effectiveness)
 - ◆ the things right (efficiency)
- ◆ Information Retrieval mechanisms are the key for automated Requirements Traceability



Discussion

- ◆ Do you think that IR mechanism can replace the work of requirements engineers?
- ◆ Do you think that automation of requirements traceability could take place in other phases than “candidate link generation”?

Requirements Tracing Process



document parsing



candidate link
generation



candidate link
evaluation



traceability
analysis



IR Methods

💧 Tf-Idf

$$w_i = \text{tf}_i(d) \cdot \text{idf}_i, \quad \text{idf}_i = \log_2\left(\frac{n}{\text{df}_i}\right)$$

$$\text{sim}(d, q) = \cos(d, q) = \frac{\sum_{i=1}^N w_i \cdot q_i}{\sqrt{\sum_{i=1}^N w_i^2 \cdot \sum_{i=1}^N q_i^2}}.$$

💧 Tf-Idf + Thesaurus

$$\langle t, t', \alpha \rangle.$$

$$\text{sim}(d, q) = \frac{\sum_{i=1}^N w_i \cdot q_i + \sum_{\langle k_i, k_j, \alpha_{ij} \rangle \in T} \alpha_{ij} (w_i \cdot q_j + w_j \cdot q_i)}{\sqrt{\sum_{i=1}^N w_i^2 \cdot \sum_{i=1}^N q_i^2}}.$$