Automating Requirements Traceability: Two Decades of Learning from KDD

Alex Dekhtyar (CSSE, Cal Poly), Jane Huffman Hayes (CS, UKy)

Learn from Knowledge Discovery in Data (KDD)

Use Machine Learning/Information Retrieval Techniques developed by researchers in KDD for automating requirements tracing

(Heh!)
Requirements Tracing is an Information Retrieval Problem

Requirements = text queries
Design elements = text documents

IR methods to establish Trace Matrix

Vectorization (numeric representation)

Similarity computation
\[ \cos(q,w) \]

Keyword weights

Measures: Precision = percentage of retrieved links that are correct
Recall = percentage of correct links that are retrieved
Lessons Learned

1. Almost 20 years of attempts (high recall, low precision)
   \[\text{Tracing} = \text{Applied KDD}\]

2. Adoption of KDD’s procedures and measures

3. Recall over Precision

4. Nobody believes our ground truth
   \[\text{Need standard procedures}\]

5. We learned to build tools
   but we need to integrate

6. Automation is not enough