Nikhitha Rajashekar
nikhita.rajashekar@rwth-aachen.de

Graph analytics approach to analyse Enterprise Architecture models

Master Thesis Proposal

Supervisor: Simon Hacks

Software Construction
RWTH AACHEN UNIVERSITY

Overview

1. Enterprise Architecture
2. Motivation
3. Research problems/Questions
4. Related work
5. Methodology
6. Summary

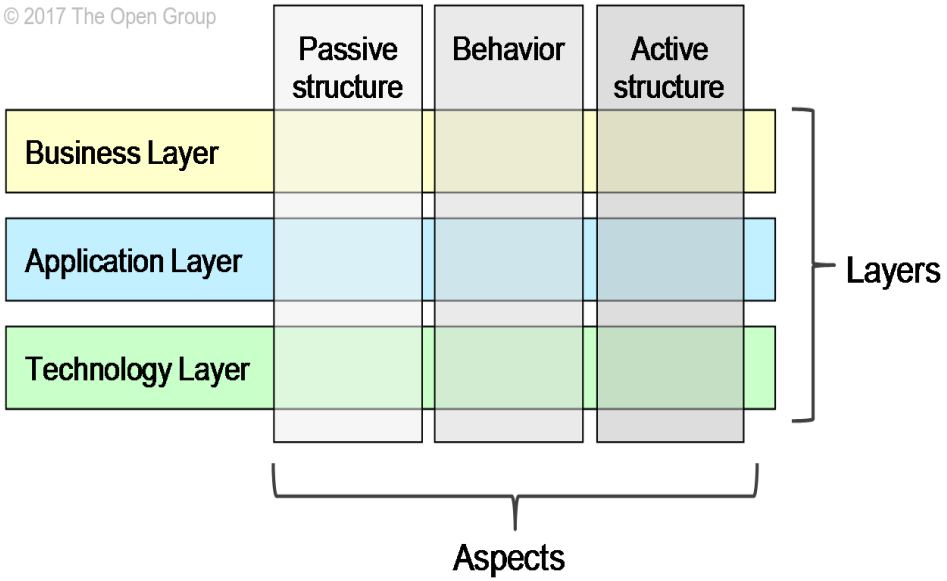
Definition


Enterprise architecture refers to the interaction between the IT components and business processes or activities of a corporation.
-
-Enterprise Architecture
Design Engineering Discipline

ArchiMate

- ArchiMate is a modeling language for describing enterprise architectures

ArchiMate Framework



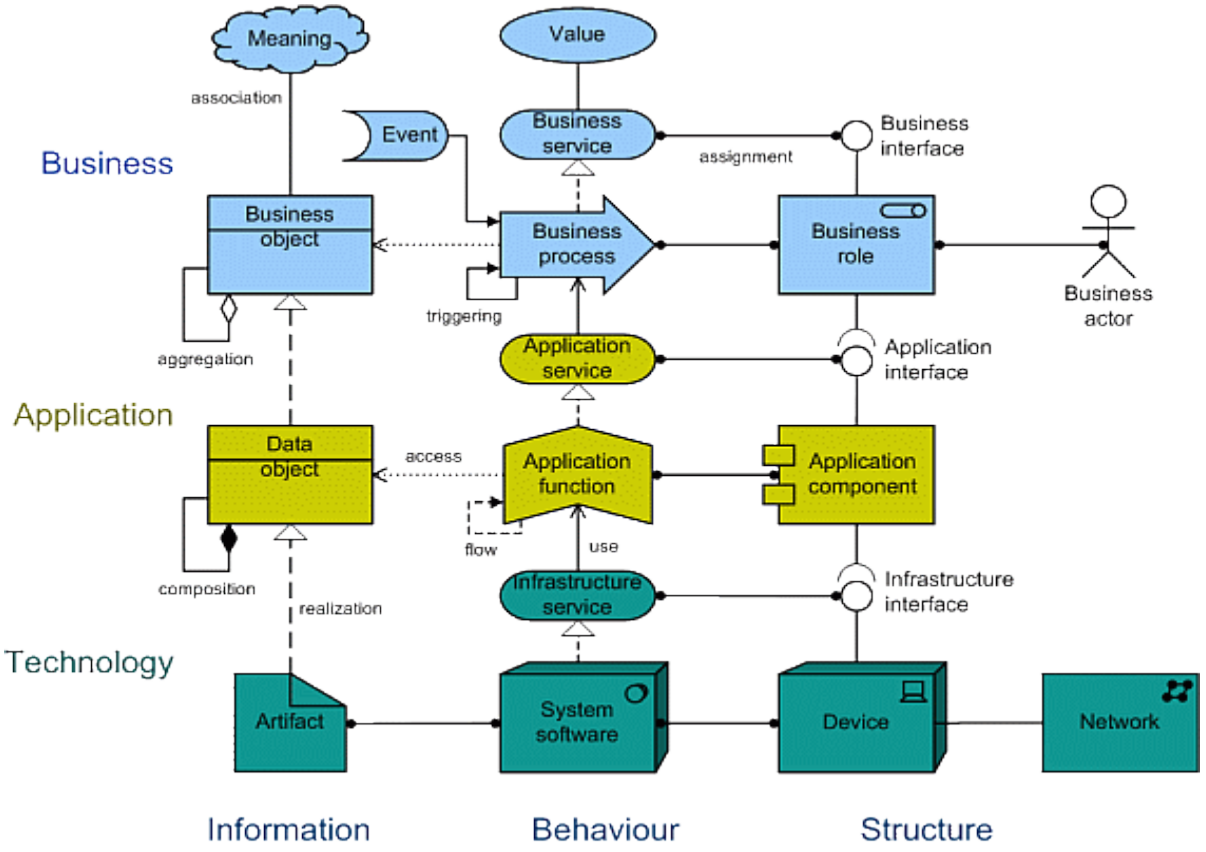
THE Open GROUP

ARCHIMATE®

SWC Software Construction | RWTH AACHEN UNIVERSITY

Archi

Archi is a open-source visual-modelling and design tool for creating ArchiMate models.



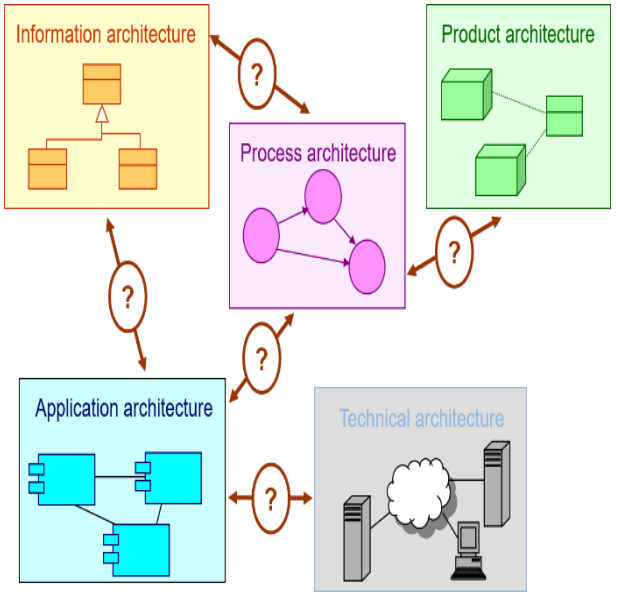
2. Motivation

Quality improvement in EAM

- Analysing EA models as a network graph.

  Vertices - Components

  Edges - Relationship
- Applying graph analytical approach and SNA to mine EA models.
- Identification of duplicate models.
- Graph based recommendation system.



3. Research problems/Questions

- Graph data.
- Presence of duplicate models.
- Ensure quality of imported project.
- Graph pattern matching.
- Selection of relevant community detection based algorithm.
- Comparing architecture models in the repository.
- Evaluation.

4. Related Work

1. Attribute based similarity:

- Similarity b/w titles, types, descriptions.
- Combining feature based similarities.

Approach	Description	Advantage	Disadvantage
Title similarity	Finds string edit distance	Detects misspelled words	Highly similar words have different meaning
Type similarity	High score for same groups	Component grouping	Restrictive matching
Description similarity	Cosine distance and TF-IDF	Semantically similar component	Only few components has description

Continued.

2. Structural similarity

- Based on node position.
- SimRank for pairwise similarity.

  Drawback: Long time to compute.

  Solution : Community detection, Random walks

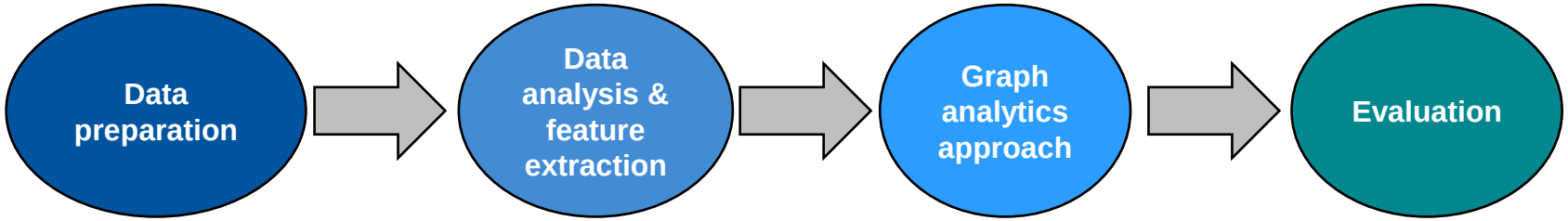
3. Association rule mining

- Conversion of architecture models from the same domain to transactions.
- Obtain the components of high interest.

5. Methodology

- Computing graph statistics.
- Extracting features(Betweenness centrality, Node degree, diameter etc.)

- Evaluating clustering algorithms by different metrics like modularity, conductance etc.
- Optimizing SimRank

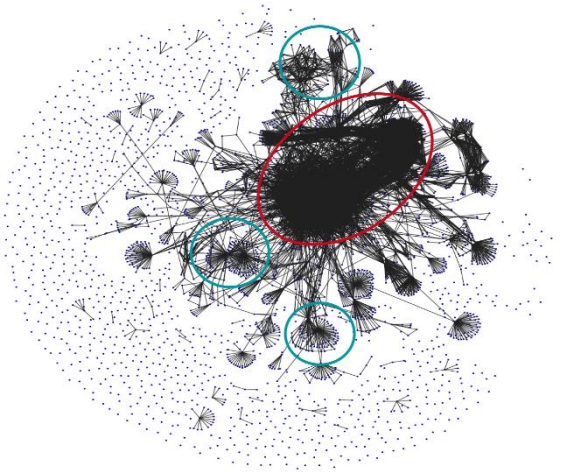


- Parsing into graph structure(using Gephi/R)

- Unsupervised learning(clustering)
- Supervised learning(Link prediction)
- Community detection algorithms
- Measuring similarity of graphs.
- Check for graph isomorphism.
- Frequent subgraph mining

Graph analytics approach

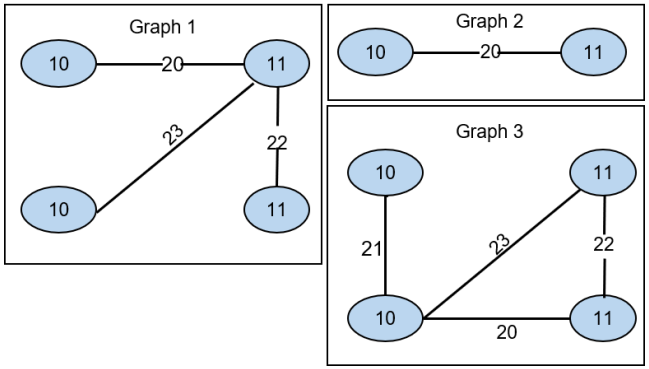
Community detection and evaluation



- Algorithms like Walktrap, edge betweenness cluster, label propagation, Louvain, fast greedy, etc.
- Evaluation metrics : Modularity, Conductance

Frequent subgraph mining

Idea: Discovery of graph structures that occur a significant number of times across a set of graphs.



Continued…

Feature extraction method

Idea: Similar graphs share certain properties like diameter, degree distribution, centrality measures.

Drawback: High similarity measure b/n two graphs that have different node set.

Measuring similarity of graphs

Idea: Graph kernel approach like Random walks.

Alternate approach

- Other similarity metrics
  - Similarity scores for vertices.
  - E.g. jaccard index, dice, inverse log weighted

- Link prediction problem???
  - Predicting link between the components.

6. Summary

- Analysing and evaluating EA models.
- Graph analytics can be applied to find duplicate models inside EA repository.
- SNA techniques in managing Enterprise architecture.
- R packages(igraph, sna, etc.) can be used to extract graph features and apply clustering algorithms.

Bibliography

1. ArchiMate® 3.0.1 Specification http://pubs.opengroup.org/architecture/archimate3-doc/toc.html
2. Aggarwal, Charu C., and Haixun Wang. "Graph data management and mining: A survey of algorithms and applications." Managing and mining graph data. Springer US, 2010. 13-68.
3. Igraph R http://igraph.org/r/doc/
4. Anthony Schoonjans. Social Network Analysis techniques in EnterpriseArchitecture Management. PhD thesis, Ghent University, 2016.
5. Koutra, Danai, et al. "Algorithms for graph similarity and subgraph matching." Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep (2011).
6. Getoor, Lise, and Christopher P. Diehl. "Link mining: a survey." Acm Sigkdd Explorations Newsletter 7.2 (2005): 3-12.

Thank you for your attention!

Any questions???