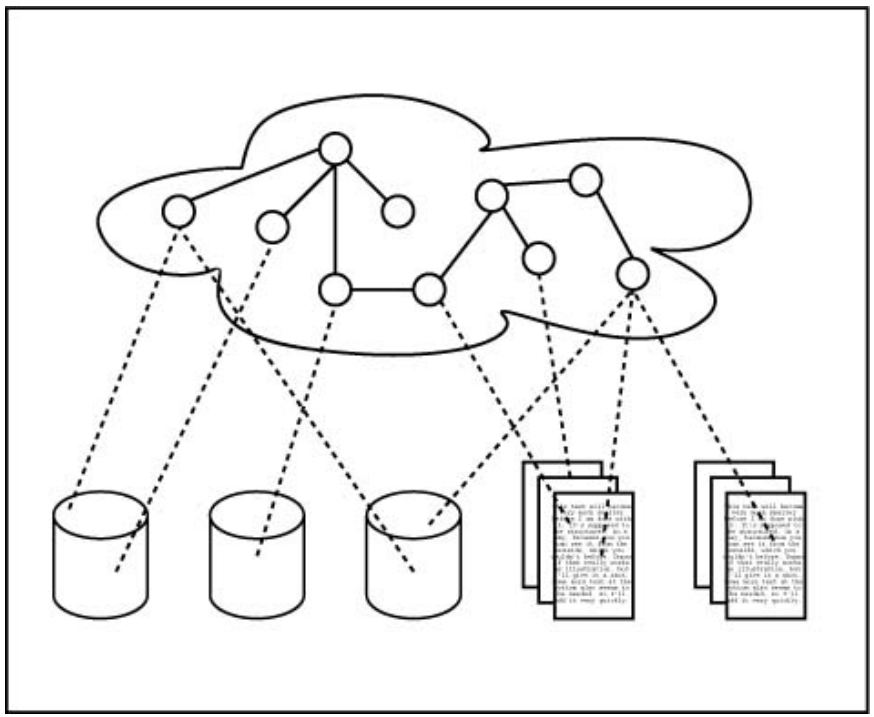
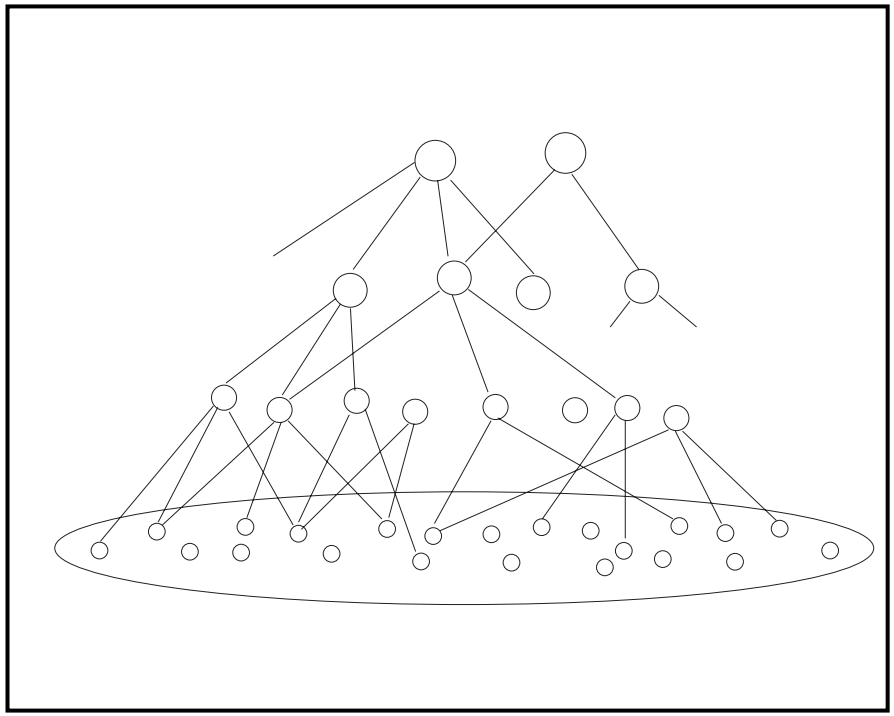
Multi-Resolution Topic Maps for Information Navigation

Azadeh Shakery

Sep. 13, 2006

Multi-Resolution Topic Map





Formal Definition

- Topic: A topic is a set of terms
- Topic Map: A topic map is a graph M = (V, E), where V is a set of topics and E is a set of topic relations.
- K-Level Multi-Resolution Topic Map: A set of k topic maps M = {M_1, …, M_k}, where M_i = (V_i, E_i) is a topic map at resolution-level i. M_1 is the lowest resolution map, while M_k is the highest resolution map.
- Reference Document Set: The reference document set of a topic t on map M_i is the subset of documents in D that the topic t represents.

Construction of an MRTM

Given a set of documents, a set of terms and levels of resolution:

1. Find all the topics in different granularities.
2. Identify the reference document set for each topic.
3. Sort the topics into different resolutions based on the sizes of their reference document sets.
4. Find the relations between topics, both in one resolution and between different resolutions.

Topic Finding Algorithm

1. Extract the representative terms of each document

2. Find the frequent itemsets of the representative terms and report them as group illustrative terms

2. Identifying Reference Documents

- Extract all the documents whose document representative terms cover the group illustrative terms

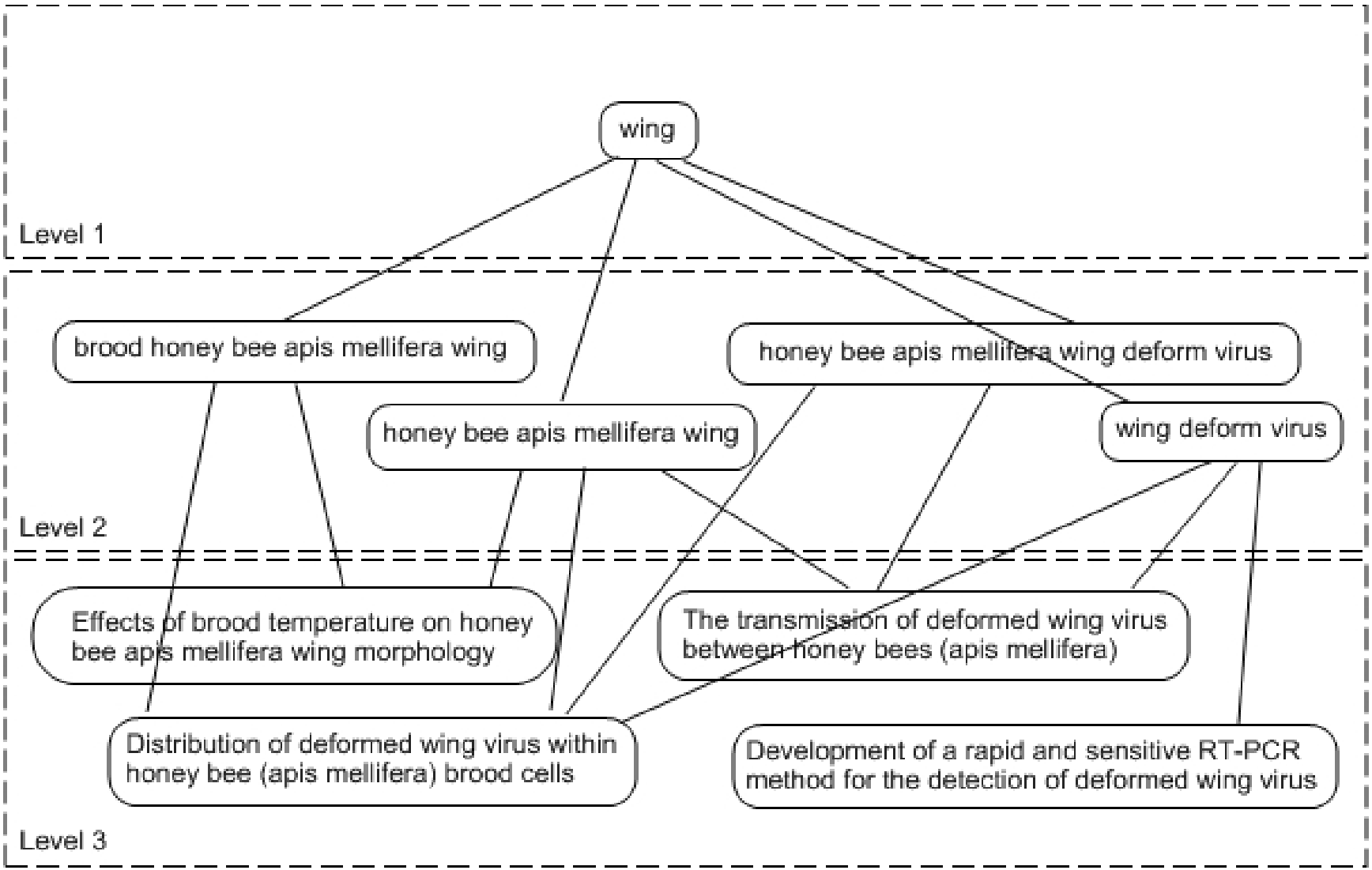
3. Grouping Topics

- Group topics based on the sizes of reference document sets
- Set topics with bigger reference document sets on high level maps and the topics with small reference document sets on lower levels
- The actual partitions of topics is application-specific
- We would like to have much fewer topics at a low resolution than at a high resolution to make the navigation efficient
- In our experiments, we have partitioned the topics such that the number of topics in different resolutions roughly form a geometric series

4. Finding Relations Between Topics

- Construct relations both between topics of the same resolution and between topics in different resolutions.
- In one resolution, two topics are related if their reference document sets overlap
- In two resolutions, a topic is the parent of another topic in a higher resolution if the reference document set of the first topic is a proper superset of the reference document set of the second topic.

Example


wing
Level 1
brood honey bee apis mellifera wing
honey bee apis mellifera wing deform virus
honey bee apis mellifera wing
wing deform virus
Level 2
Effects of brood temperature on honey bee apis mellifera wing morphology
The transmission of deformed wing virus between honey bees (apis mellifera)
Distribution of deformed wing virus within honey bee (apis mellifera) brood cells
Development of a rapid and sensitive RT-PCR method for the detection of deformed wing virus
Level 3

Demo