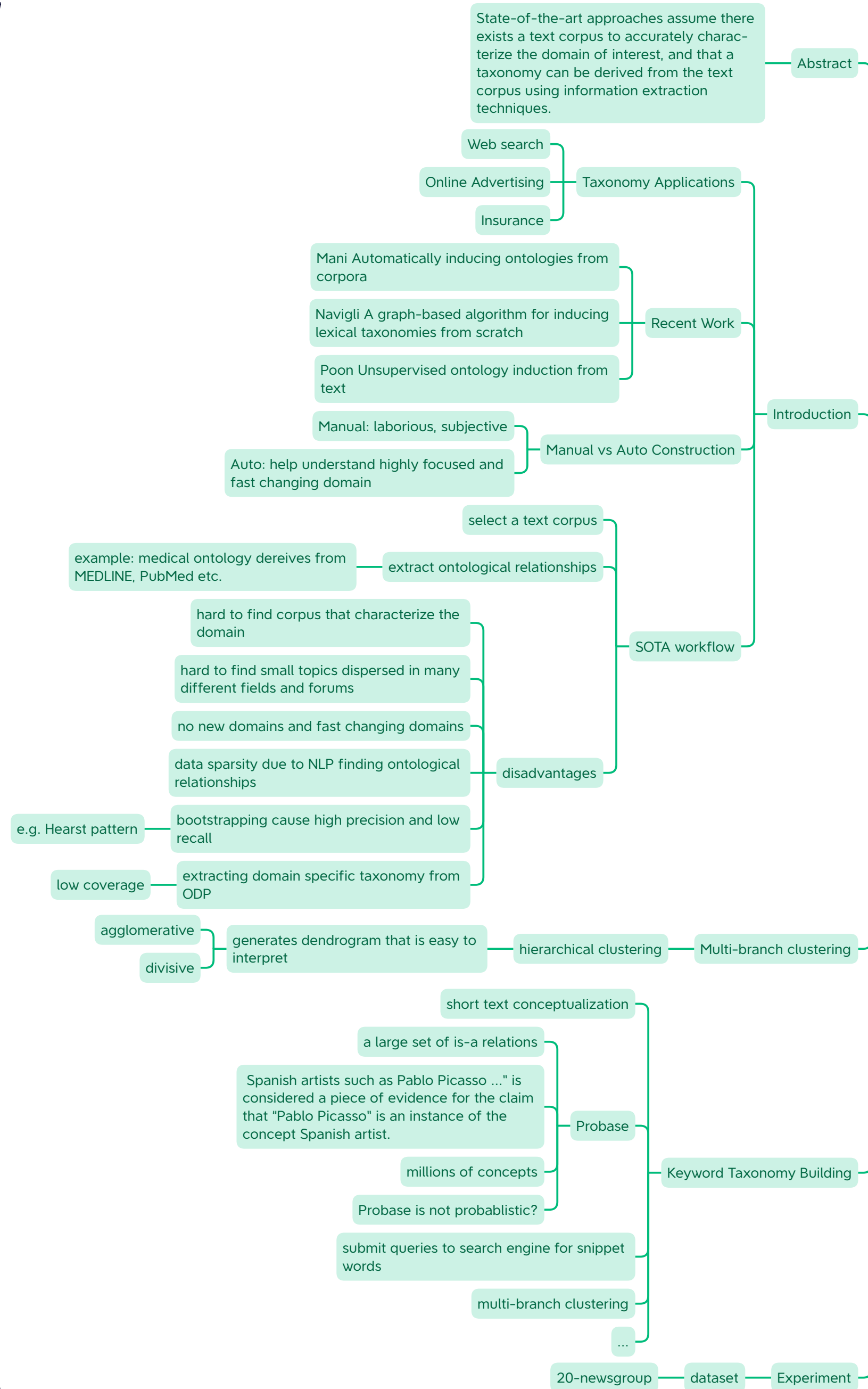


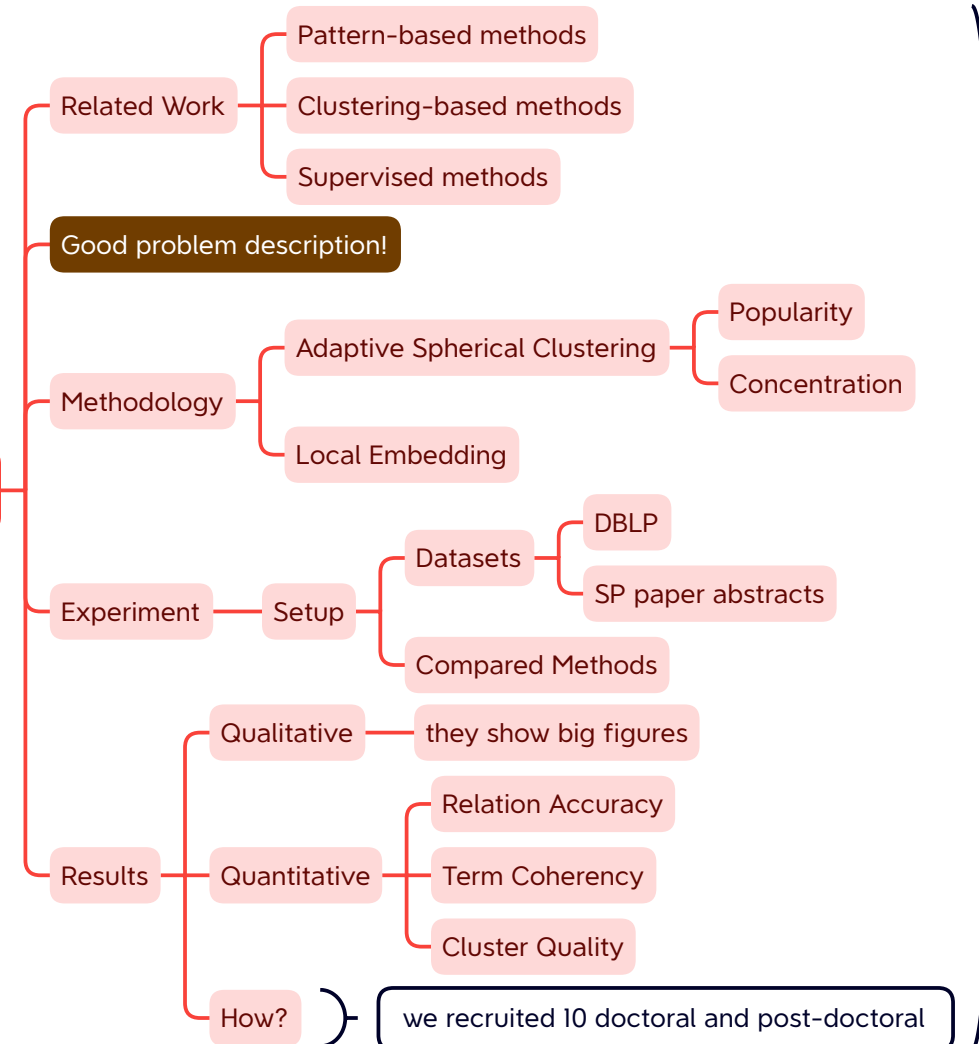
Construction Papers

2012 Automatic Taxonomy Construction from Keywords



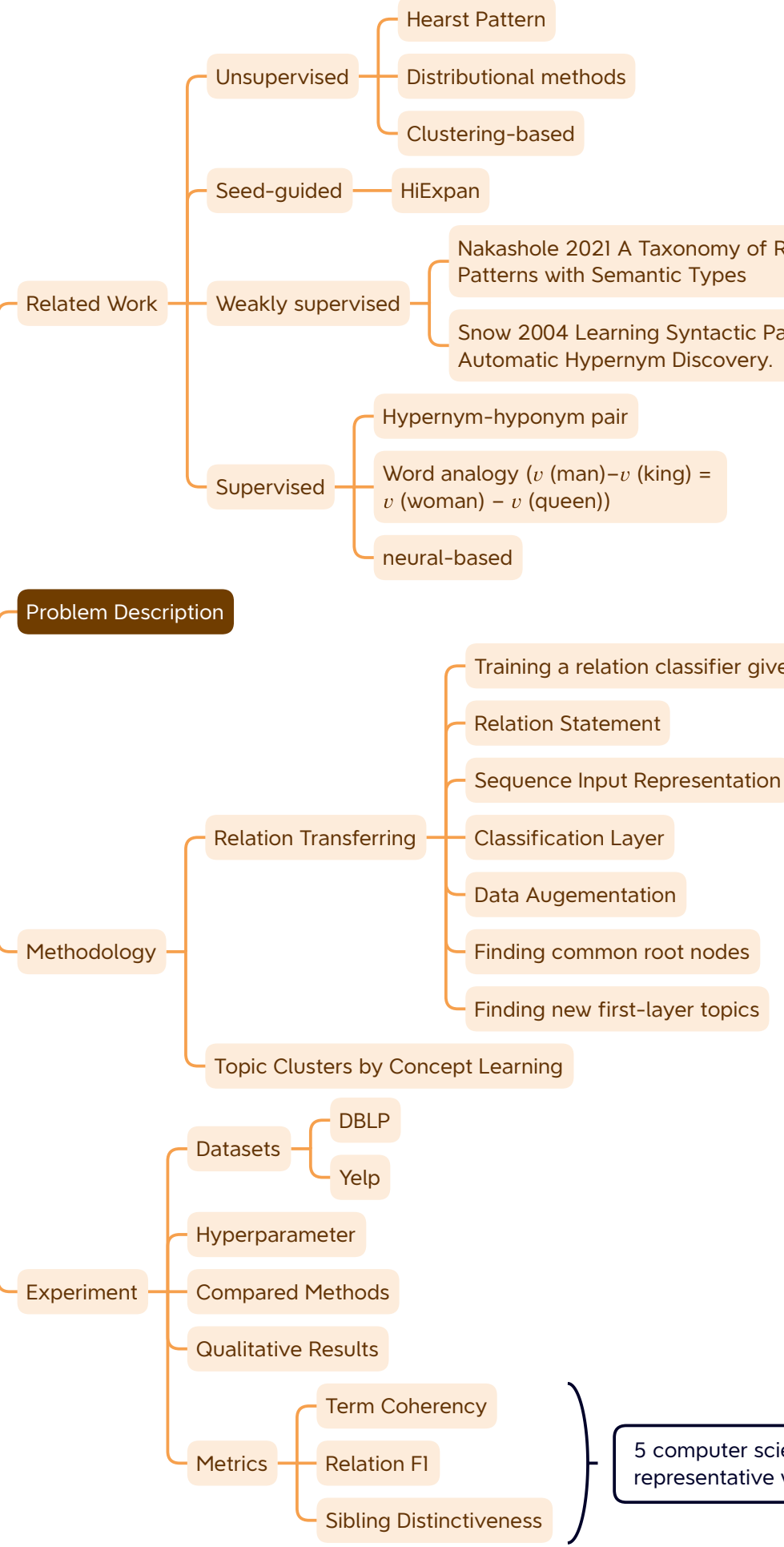
In this paper, we present an approach that can automatically derive a domain-dependent taxonomy from a set of keyword phrases by leveraging both a general knowledgebase and keyword search. We first deduce concepts with the technique of conceptualization and extract context information from a search engine, and then induce the new taxonomy using a Bayesian rose tree. We provide three nearest-neighborhood-based methods to speed up the original Bayesian rose tree algorithm. Particularly, the Spilltree-based algorithm reduces the time and memory cost significantly. We also conducted a set of experiments to demonstrate the effectiveness and efficiency of the proposed algorithms.

TaxoGen 2018



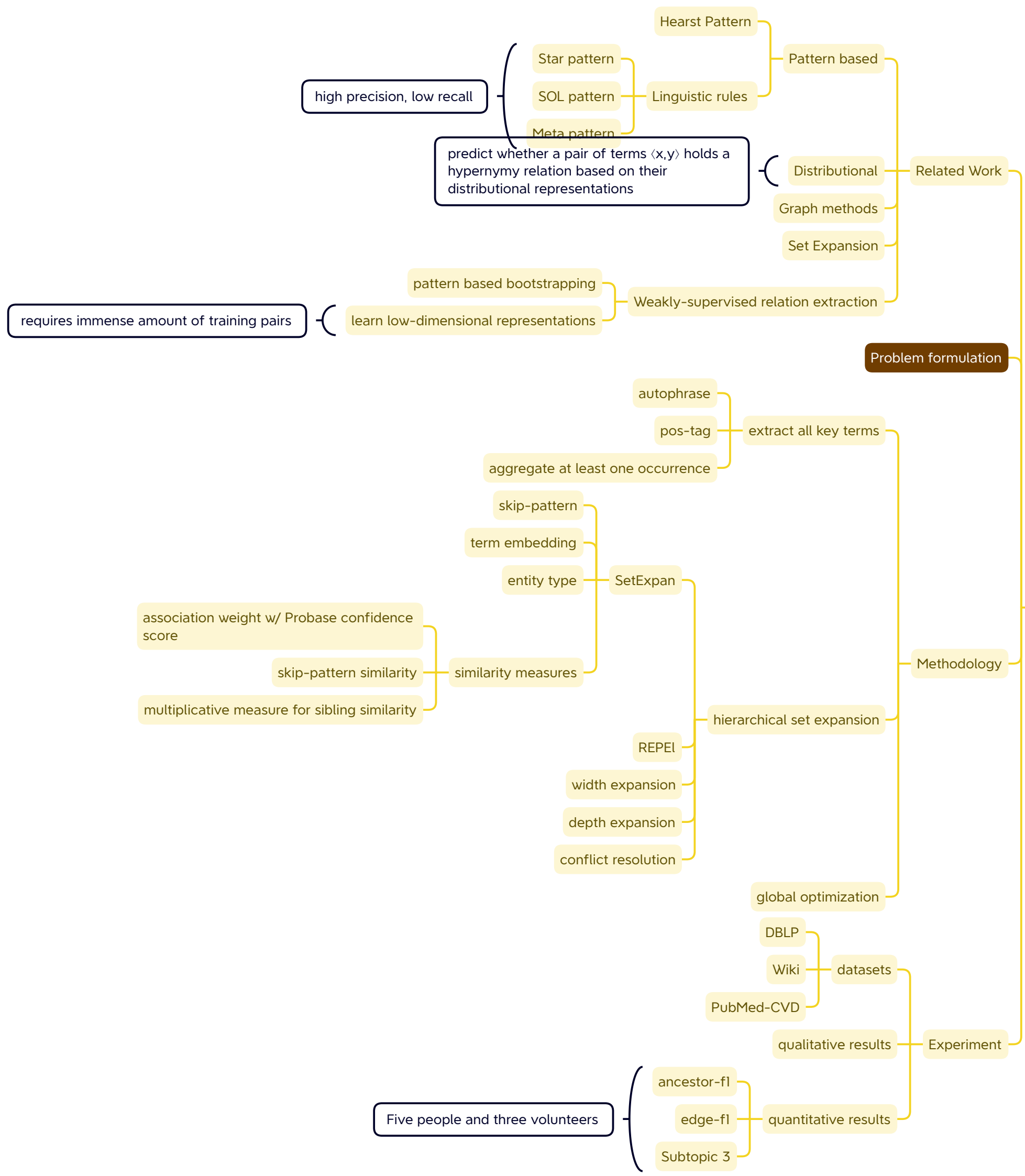
We studied the problem of constructing topic taxonomies from a given text corpus. Our proposed method TaxoGen relies on term embedding and spherical clustering to construct a topic taxonomy in a recursive way. It consists of an adaptive clustering module that allocates terms to proper levels when splitting a coarse topic, as well as a local embedding module that learns term embeddings to maintain strong discriminative power at lower levels. In our experiments, we have demonstrated that both two modules are useful in improving the quality of the resultant taxonomy, which renders TaxoGen advantages over existing methods for building topic taxonomies. One limitation of the current version of TaxoGen is that it requires a pre-specified number of clusters when splitting a coarse topic into fine-grained ones. In the future, it is interesting to extend TaxoGen to allow it to automatically determine the optimal number of children for each parent topic in the construction process.

CoRel 2020



In this paper we explore the problem of seed-guided topical taxonomy construction. Our proposed framework CoRel completes the taxonomy structure by a relation transferring module and enriches the semantics of concept nodes by a concept learning module. The relation transferring module learns the user-interested relation preserved in seed parent-child pairs, then transfers it along multiple paths to expand the taxonomy in width and depth. The concept learning module finds discriminative topical clusters for each concept in the process of jointly embedding concepts and words. Extensive experiments show that both modules work effectively in generating a high-quality topical taxonomy based on user-given seeds.

HiExpan 2018



In this paper, we introduce a new research problem task-guided taxonomy construction and propose a novel expansion-based framework HiExpan for solving it. HiExpan views all children under a taxonomy node as a coherent set and builds the taxonomy by recursively expanding these sets. Furthermore, HiExpan incorporates a weakly-supervised relation extraction module to infer parent-child relation and adjusts the taxonomy tree by optimizing its global structure. Experimental results on three public datasets corroborate the effectiveness of HiExpan.