Chapter 11

Related Work

Previous research efforts on generating IE rules from texts with templates include AutoSlog-TS [49], CRYSTAL [58], PALKA [25], LIEP [19], RAPIER [9] and WHISK [59]. AutoSlog-TS creates dictionaries of extraction patterns using syntactic heuristics from untagged texts. PALKA generates Frame-Phrasal pattern structures, but it did not prove to be noise tolerant. CRYSTAL employed inductive learning techniques to generate *concept nodes* used by the BADGER sentence analyzer. LIEP used a set of heuristics to learn dictionaries of extraction patterns. Another approach is the learning of pattern-match rules that specify pre-filler, filler and post-filler patterns. This approach was taken by RAPIER [9]. RAPIER learns pattern-match rules. RAPIER's learning method is strongly inspired by ILP techniques. WHISK is a learner that generates superficial extraction patterns from texts with tags, such as HTML tags.

In our approach, the hierarchically sorted ILP system RHB$^+$ is independent of natural language analysis. This point differentiates our approach from the others. That is, learned logic programs may have several atomic formulae in their bodies. Our approach is different from simple generalization of a single frame of semantic representations. That is, we believe the information in several case frames can be complementary to achieving an appropriate generality of an IE rule.

Sasaki [52] reported results of preliminary experiments on learning IE rules to extract information from twenty articles with the preliminary version of the ILP system, which was capable of handling most

features of ψ-terms. The hypothesis language of the preliminary version of the learner has more representation power than that of RHB$^+$, but this power comes at the expense of learning speed.