# Forward

The *Workshop on Statistically-Based NLP Techniques* held at the Tenth National Conference on Artificial Intelligence in San Jose was the first AAAI workshop to address a growing interest among members of the NL research community in statistically-based techniques. The papers from this workshop are representative of the diversity of approaches now being explored by investigators striving to integrate statistically-based techniques with the knowledge-based techniques that have traditionally dominated NLP research initiatives.

The following is from the call for participation for the workshop:

> Interest in statistically-based NLP techniques has grown considerably over the last five years, partly because of disenchantment with the rate of technological progress in developing NLP systems within a strictly "knowledge-based" framework. Such systems have suffered from three chronic problems. First, their reliance upon domain restrictions tends to result in a lack of robustness when confronted with gaps in coverage. Second, because domain knowledge is handcoded in such systems, extending them to support new domains tends to be a laborious process. Third, such systems generally must be maintained by developers, not by users.
>
> There are, of course, good reasons why researchers have developed NLP systems within a knowledge-based framework—some information is very difficult to capture and represent by statistical means. Rather than completely abandoning a knowledge-based framework, researchers have begun to develop hybrid systems in which an effort is made to maximize the potential of statistically-based and knowledge-based techniques.
>
> With the growing interest in statistically-based techniques, it is time for a forum on their use in NLP applications. What components of an NLP system can benefit from such techniques? What tradeoffs exist in using statistical techniques, and in combining them with handcrafted knowledge? Are there interesting interactions that arise when more than one such technique is used? And finally, is there evidence that a given technique is capable of supporting large-scale applications—for example, is it reasonable to expect grammar induction systems to be capable of generating broad-coverage grammars capable of supporting large-scale data extraction applications, and if so, are there any special benefits of using such an approach in a large-scale system?
>
> The objective of this workshop is to establish the capabilities of existing statistically-based NLP techniques, and to envision how they may be improved. Discussions of significant success stories and interesting failures in efforts to employ such techniques within large-scale NLP applications will be emphasized. Reports on the use of statistically-based methods in syntactic and semantic analysis will be especially encouraged, along with reports on efforts to automate the acquisition of linguistic knowledge from large text and spoken language corpora.