MITRE's Qanda at TREC-12

John D. Burger
The MITRE Corporation
john@mitre.org

Introduction

Qanda is MITRE's TREC-style question answering system. This year, we were able to apply only a small effort to the TREC QA activity, approximately one person-month. As well as some general improvements in Qanda's processing, we made some simple attempts to handle definition and list answers.

1. TREC-12 system description

Catalyst architecture

Qanda uses a general architecture for human language technology called *Catalyst*, (Burger & Mardis 2002, Nyberg et al., to appear). Developed at MITRE for the DARPA TIDES program, the Catalyst architecture is specifically designed for fast processing and for combining the strengths of Information Retrieval and Natural Language Processing into a single framework. Catalyst uses a dataflow architecture in which standoff annotations are passed from one component to another, with the components connected in arbitrary (acyclic) topologies. The use of standoff annotation permits components to be optimized for just those pieces of information they require for their processing.

Major system components

Qanda has a by now familiar architecture—questions are analyzed for expected answer types, documents are retrieved using an IR system and are then processed by various taggers to find entities of the expected types in contexts that match the question. Below we describe each of the major components in turn.