



THE UNIVERSITY OF ARIZONA
COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Final Ph.D. Examination

Of

Enrique Noriega

**“Assembling Information
from Big Corpora by Focusing
Machine Reading”**

Committee: Dr. Clayton Morrison (Advisor)
Dr. Mihai Surdeanu
Dr. Peter Jansen

Wednesday, December 16th, 2020
10:00am

Join at <https://arizona.zoom.us/j/86169523455>

Password: 241294

Abstract:

We propose a framework to teach an automated agent to learn how to search for multi-hop paths of relations between entities in large corpora. The method learns a policy for directing existing information retrieval and machine reading resources to focus on relevant regions of a corpus.

The approach formulates the learning problem as a Markov decision process with a state representation that encodes the dynamics of the search process and a reward structure that minimizes the number of documents that must be processed while still finding multi-hop paths.

We implement the framework with reinforcement learning and evaluate it on an open-domain dataset of search problems derived from a subset of English Wikipedia using a policy gradient actor-critic algorithm and a domain-specific dataset of search problems in the biomedical domain using a temporal-difference learning algorithm.

We show that deploying the focused reading framework with reinforcement learning finds policies that retrieve more multi-hop paths while processing fewer documents compared to several strong deterministic baseline implementations.