

**SEARCHING FOR FRAGMENTS IN LARGE
SEMANTIC DATABASES**

By

Bridget M. Gutting

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: **COMPUTER SCIENCE**

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Approved:

Dr. Mark Goldberg, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York

April 2012
(For Graduation May 2012)

ABSTRACT

This thesis presents a novel indexing and searching schemes for semantic graphs which are based on the i -degrees of graph nodes. An i -degree of a node is a length t vector which calculates the number of nodes of each type at a distance i from the central node. The i -degrees use relationship and type information for specifying nodes rather than text labels. The goal is to be able to search for a small fragment graph within a larger semantic graph (database) where the fragment is a representation of information a research has gathered. Fragments may contain information which is incomplete, inaccurate, or unreliable resulting in the difficulty of identifying its position within the database. For example, the label of a “person” node may be unreliable, the real name is “Bridget” not “Brittany,” so searching by the label would likely produce irrelevant results. Since this is problem is closely related to isomorphic subgraph identification which is an NP-hard problem, the algorithms presented are efficient heuristics. Results are presented from Wikipedia and Random graphs.