

A Generalized Flow-Based Method for Analysis of Implicit Relationships on Wikipedia

ABSTRACT

We focus on measuring relationships between pairs of objects in Wikipedia whose pages can be regarded as individual objects. Two kinds of relationships between two objects exist: in Wikipedia, an explicit relationship is represented by a single link between the two pages for the objects, and an implicit relationship is represented by a link structure containing the two pages. Some of the previously proposed methods for measuring relationships are cohesion-based methods, which underestimate objects having high degrees, although such objects could be important in constituting relationships in Wikipedia. The other methods are inadequate for measuring implicit relationships because they use only one or two of the following three important factors: distance, connectivity, and cocitation. We propose a new method using a generalized maximum flow which reflects all the three factors and does not underestimate objects having high degree. We confirm through experiments that our method can measure the strength of a relationship more appropriately than these previously proposed methods do. Another remarkable aspect of our method is mining elucidatory objects, that is, objects constituting a relationship. We explain that mining elucidatory objects would open a novel way to deeply understand a relationship.

EXISTING SYSTEM:

In Existing several semantic search engines have been used for searching relationships between two objects, using a semantic knowledge base extracted from web or Wikipedia. However, the semantics in these knowledge bases, such as “isCalled,” “type” and “subClassOf,” are mainly used to construct an

ontology for objects. Such semantic knowledge bases are still far from covering relationships existing in Wikipedia, such as “Gulf of Mexico” is a major “petroleum” producer. We do not utilize the semantic knowledge bases for measuring relationships in this paper.

PROPOSED SYSTEM:

We propose a new method for measuring a relationship on Wikipedia by reflecting all the three concepts: distance, connectivity, and cocitation. We measure relationships rather than similarities.

1. A detailed and methodical survey of related work for measuring relationships or similarities.
2. A new method using generalized maximum flow for measuring the strength of a relationship between two objects on Wikipedia, which reflects the three concepts: distance, connectivity, and cocitation.
3. Experiments on Wikipedia showing that our method is the most appropriate one .
4. Case studies of mining elucidatory objects for deeply understanding a relationship .

MODULE DESCRIPTION:

Number of Modules

After careful analysis the system has been identified to have the following modules:

- 1. Link Analysis Module.**
- 2. Generalized Flow Based Module.**
- 3. Wikipedia Mining Module.**
- 4. Relationship on Wikipedia Module.**

1. Link Analysis Module:

Two kinds of relationships between two objects exist: in Wikipedia, an explicit relationship is represented by a single link between the two pages for the objects, and an implicit relationship is represented by a link structure containing the two pages. A user also might desire to discover a relationship between two objects. For example, a user might desire to know which countries are strongly related to petroleum, or to know why one country has a stronger relationship to petroleum than another country. Typical keyword search engines can neither measure nor explain the strength of a relationship. The main issue for measuring relationships arises from the fact that two kinds of relationships: “explicit relationships” and “implicit relationships.” In Wikipedia, an explicit relationship is represented by a link. An implicit relationship is represented by multiple links and pages. For example, an implicit relationship between petroleum and the USA might be represented by links.

2. Generalized Flow Based Module:

The three concepts, distance, connectivity, and cocitation, are important concepts for measuring relationships; cohesion-based methods underestimate popular objects, although popular objects might be important for relationships in Wikipedia. Our method can mine elucidatory objects constituting a relationship by outputting paths contributing to the generalized maximum flow, that is, paths along which a large amount of flow is sent. We propose a generalized maximum flow-based method which reflects all the three concepts and does not underestimate popular objects, in order to measure relationships on Wikipedia appropriately.

3. Wikipedia Mining Module:

Searching webpages containing a keyword has grown in this decade, while knowledge search has recently been researched to obtain knowledge of a single object and relationships between multiple objects, such as humans, places or events. Searching knowledge of objects using Wikipedia is one of the hottest topics in the field of knowledge search. In Wikipedia, the knowledge of an object is gathered in a single page updated constantly by a number of volunteers. Wikipedia also covers objects in a number of categories, such as people, science, geography, politic, and history. Therefore, searching Wikipedia is usually a better choice for a user to obtain knowledge of a single object than typical search engines.

4. Relationship On Wikipedia Module:

We propose a new method for measuring a relationship on Wikipedia by reflecting all the three concepts: distance, connectivity, and cocitation. We propose a new method for measuring the strength of a relationship using the generalized maximum flow. The value of flow f is defined as the total amount of f arriving at destination t . To measure the strength of a relationship from object s to object t , we use the value of a generalized maximum flow emanating from s as the source into t as the destination; a larger value signifies a stronger relationship. We regard the vertices in the paths composing the generalized maximum flow as the objects constituting the relationship. We qualitatively ascertain the claim that our method can reflect the three representative concepts.

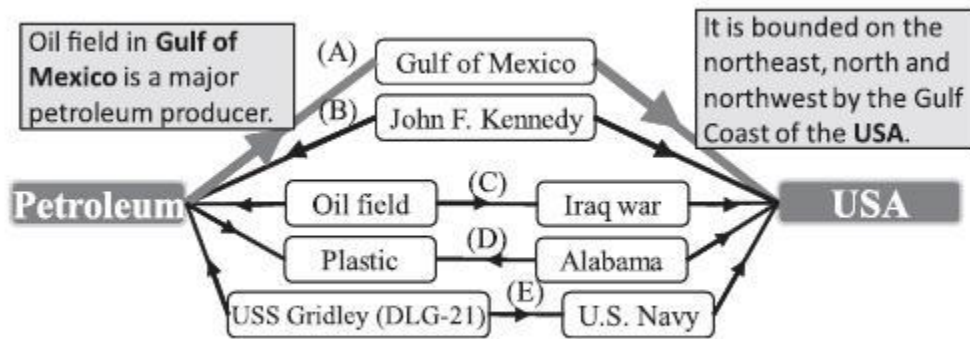
SOFTWARE REQUIREMENTS:

Operating System	: Windows
Technology	: Java and J2EE
Web Technologies	: Html, JavaScript, CSS
IDE	: My Eclipse
Web Server	: Tomcat
Tool kit	: Android Phone
Database	: My SQL
Java Version	: J2SDK1.5

HARDWARE REQUIREMENTS:

Hardware	: Pentium
Speed	: 1.1 GHz
RAM	: 1GB
Hard Disk	: 20 GB
Floppy Drive	: 1.44 MB
Key Board	: Standard Windows Keyboard
Mouse	: Two or Three Button Mouse
Monitor	: SVGA

Flow Diagram:



Explaining the relationship between Petroleum and the USA.