



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

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 co citation. 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



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

constituting a relationship. We explain that mining elucidatory objects would open a novel way to deeply understand a relationship.

EXISTING SYSTEM:

Several methods have been proposed for measuring the strength of a relationship between two objects on an information network (V, E) , a directed graph where V is a set of objects; an edge $(u, v) \in E$ exists if and only if object $u \in V$ has an explicit relationship to $v \in V$. We can define a Wikipedia information network whose vertices are pages of Wikipedia and whose edges are links between pages. Previously proposed methods then can be applied to Wikipedia by using a Wikipedia information network. The Concept of “cohesion,” exists for measuring the strength of an implicit relationship. CFEC proposed by Koren et al. [1] and PFIBF proposed by Nakayama et al. is based on cohesion. We do not adopt the idea of cohesion based methods, because they always punish objects having high degrees although such objects could be important to some relationships in Wikipedia. Other previously proposed methods use only one or two of the three representative concepts for measuring a relationship: distance, connectivity, and cocitation, although all the concepts are important factors for implicit relationships. Using all the three concepts together would be appropriate for measuring an implicit relationship and mining elucidatory objects.



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

DISADVANTAGES OF EXISTING SYSTEM:

- It is difficult for the user to discover an implicit relationship and elucidatory objects without investigating a number of pages and links.
- Therefore, it is an interesting problem to measure and explain the strength of an implicit relationship between two objects in Wikipedia.

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. As discussed in relationship is a more general concept than similarity. For example, it is hard to say petroleum is similar to USA, but a relationship exists between petroleum and the USA. Our method uses a “generalized maximum flow” on an information network to compute the strength of a relationship from object s to object t using the value of the flow whose source is s and destination is t . It introduces a gain for every edge on the network. The value of a flow sent along an edge is multiplied by the gain of the edge. Assignment of the gain to each edge is important for measuring a relationship using a generalized maximum flow. We propose a heuristic gain function utilizing the category structure in Wikipedia. We confirm through experiments that the gain function is sufficient to measure relationships appropriately.



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

ADVANTAGES OF PROPOSED SYSTEM:

- Compute the strength of the relationship between a source object and each of its destination objects, and rank the destination objects by the strength.
- Assignment of the gain to each edge is important for measuring a relationship using a generalized maximum flow.
- Experiments on Wikipedia showing that our method is the most appropriate one

MODULES:

- ⊗ Link Analysis Module.
- ⊗ Generalized Flow Based Module.
- ⊗ Wikipedia Mining Module.
- ⊗ Ranking Module.

MODULES DESCRIPTION:

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



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

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.

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 under estimates popular objects, in order to measure relationships on Wikipedia appropriately.



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry

Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

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.

Ranking Module:

We propose a new method for measuring the strength of a relationship using the generalized maximum flow. In addition we propose a ranking module, where the A good evaluation of methods measuring relationships always requires human subjects ranking.



JP INFOTECH

SOFTWARE DEVELOPMENT & RESEARCH DIVISION

www.jpinfotech.org

(0)9952649690

Chennai Office : JP INFOTECH, Old No.31,
New No. 86, 1st Floor , 1st Avenue , Ashok
Pillar , Chennai - 83.

Landmark : Next to Kotak Mahendra Bank.

Pondicherry Office : JP INFOTECH , #45 ,
Kamaraj Salai, Thattanchavady, Puducherry
Landmark : Next to VVP Nagar Arch.

jpinfotechprojects@gmail.com

SYSTEM CONFIGURATION:-

HARDWARE CONFIGURATION:-

- ✓ Processor - Pentium –IV
- ✓ Speed - 1.1 Ghz
- ✓ RAM - 256 MB(min)
- ✓ Hard Disk - 20 GB
- ✓ Key Board - Standard Windows Keyboard
- ✓ Mouse - Two or Three Button Mouse
- ✓ Monitor - SVGA

SOFTWARE CONFIGURATION:-

- ✓ Operating System : Windows XP
- ✓ Programming Language : JAVA/J2EE.
- ✓ Java Version : JDK 1.6 & above.
- ✓ Database : MYSQL