Ranked Search on Data Graphs

Ramakrishna Varadarajan

Thesis Proposal

FLORIDA INTERNATIONAL UNIVERSITY,
School of Computing and Information Sciences,
Miami.

Acknowledgments to the Committee

Advisor: Professor: Vagelis Hristidis

Committee Members:

Professor: Shu-Ching Chen

Professor: Tao Li

Professor: Raju Rangaswami

Professor: Kaushik Dutta, FIU College of Business.

Roadmap

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
- Ongoing & Future Work

Roadmap

- **Problem Statement & Motivation**
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
- Ongoing & Future Work

Problem Statement & Motivation

- Graph-structured databases – becoming a commonplace.
- Need for Efficient & High Quality search & retrieval.
- Common Graph Models
  - Web
    - Nodes – Pages
    - Edges – Hyperlinks
  - Relational Database
    - Nodes – Tuples
    - Edges – Primary/Foreign key relationships
  - XML
    - Nodes – XML elements
    - Edges – Intra-document links (IDREFs), Inter-document links (Xlinks)

Problem Statement & Motivation

- ***Keyword Search*** – most effective & dominant information discovery method.
- Success of search engines confirm this.
- Key Advantages:
  - ❑ Simplicity (ease of use).
  - ❑ Query interface is flexible.
  - ❑ No prior knowledge about structure of underlying data.
  - ❑ Queries can be imprecise
- Recently applied over Structured (databases) & Semi-structured Data(XML).

Roadmap

- Problem Statement & Motivation
- **State of Art Graph Search Methods**
- Data Model
- Related Work
- Preliminary Work
- Ongoing & Future Work

State of Art Graph Search Methods

- ***Keyword Proximity Search***
  - Input: Data Graph *DG*, Keyword Query *Q*, Ranking Function *f*, Top-*k*.
  - Output: *k* subgraphs of *DG* with smallest (or largest) scores such that each of the subgraph is
    - A Tree.
    - Total (contains all the keywords).
    - Minimal (non-redundant).

- **Applications**:
  - Web ("Information Unit" paper [WWW02]).
  - Database (DBXplorer [ICDE02], BANKS [ICDE02], DISCOVER [VLDB02], IRStyle [VLDB03],GoldMan [VLDB98]).
  - XML (Xkeyword [ICDE03], Xsearch [VLDB03]).

State of Art Graph Search Methods

- ***Authority Flow-Based Search***
  - ❖ Input: Data Graph *DG*, Keyword Query *Q*, Top-*k*.
  - ❖ Output: *k* nodes of *DG* of highest global importance and relevance to the query. Rankings are
    - ❑ Primarily based on underlying link-structure.
    - ❑ Secondarily based on content.

- **Applications**:
  - ➢ Web(PageRank [WWW98], Topic-Sensitive PageRank [WWW02], Scaling Personalized Web Search [WWW03]).
  - ➢ Database (ObjectRank[VLDB04]).
  - ➢ XML (XRANK[SIGMOD03]).

Roadmap

- Problem Statement & Motivation
- State of Art Graph Search Methods
- **Data Model**
- Related Work
- Preliminary Work
- Ongoing & Future Work

Data Model

- ***Web Graph (directed, unweighted)***
  - Web Pages (nodes) & Hyperlinks (edges)

Page 1 http://www.fiu.edu/orgs/admiss/index.htm

(**v1**) Admission Office - Florida International University.
(**v2**) …..
(**v3**) Florida International University, a member of the State University System of Florida, is a fully accredited comprehensive, multi-campus urban **research** institution located in Miami, Florida (more).
(**v4**) Open House, Latest **Scholarships**, The Honors College.
(**v5**) ……
(**v6**) ……
(**v7**) ……

Page 4 http://www.fiu.edu/docs/brief_history4.htm

(**v1**) Florida International University: A Brief History.
(**v2**) ACHIEVING THE UNIVERSITY'S VISION 1986 TO THE PRESENT.
(**v3**) …..
(**v4**) ……
(**v5**) ……
(**v6**) ……
(**v7**) ……
(**v8**) ……
(**v9**) In early 1996, in order to secure resources for faculty endowed chairs, **scholarships** and facilities, the University launched its first major capital campaign, the $65 million Campaign for FIU…… This goal that was achieved in early 2001, more than one year ahead ahead of schedule.
(**v10**) ……
(**v11**) ……
(**v12**) Some 30 years after opening its doors, FIU is within reach of attaining its foremost goal - to become one of the nation's top, urban, public **research** universities……...
(**v13**) …..

Page 2 http://www.fiu.edu/orgs/finaid/programs/programs.htm

(**v1**) Financial Aid Office.
(**v2**) International Students.
(**v3**) …..
(**v4**) ……
(**v5**) **Graduate** Assistance - **Graduate** students pursuing a master or doctoral degree may qualify for assistantships, fellowships and other awards offered through individual academic units…….. To inquire regarding additional **graduate** information contact the Division of **Graduate** Studies Office.
(**v6**) ……
(**v7**) ……
(**v8**) ……
(**v9**) ……
(**v10**) ……
(**v11**) Institutional Programs Academic Merit Assistance The University's commitment to academic excellence…….. Additional awards for outstanding high school seniors include the Presidential Scholars, Academic Excellence Scholars, Valedictorian and Salutatorian **Scholarships**……..
(**v12**) ……
(**v13**) …..

Page 3 http://www.fiu.edu/index.html

(**v1**) Florida International University: Miami's Public **Research** University.
(**v2**) …..
(**v3**) ADMISSIONS: **GRADUATE** - ARTS & CULTURE - ATHLETICS - BISCAYNE BAY CAMPUS - ……... – PHONEBOOK -REGISTRATION - STUDENT GOVERNMENT - UNIVERSITY **GRADUATE** SCHOOL - VALUES STATEMENT – WEB MAIL- SEARCH FIU - ADVANCED SEARCH.
(**v4**) ……
(**v5**) ……
(**v6**) ……

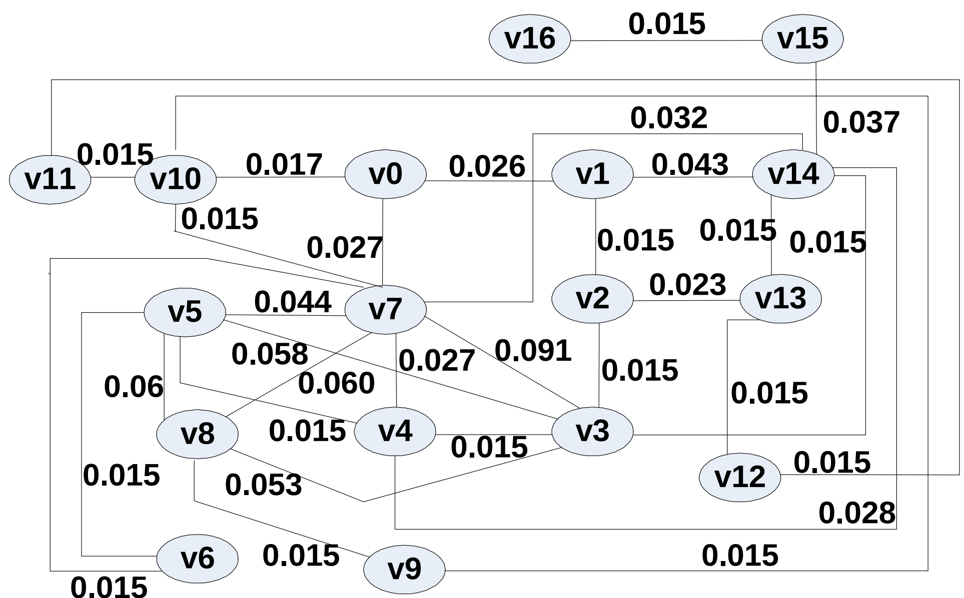
LEGEND — Node in Web Graph (Web Page); Hyperlinks

Data Model

Sample Document

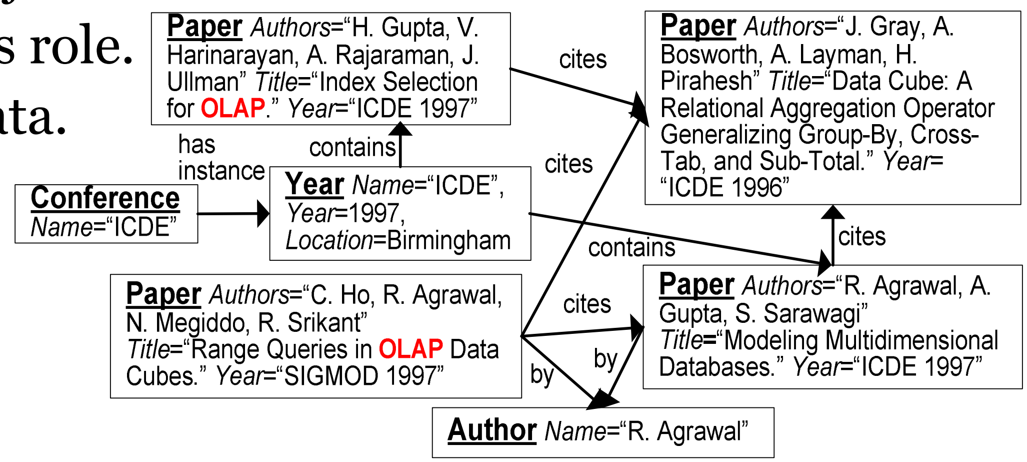
(v0) Brain chip offers hope for paralyzed
(v1) A team of neuroscientists have successfully implanted a **chip** into the **brain** of a quadriplegic man, allowing him to control a computer.
(v2) …
(v3) …
(v4) …
(v5) BrainGate offers the possibility of hitherto unimaginable levels of independence for the severely disabled.
(v6) …
(v7) …
(v8) …
(v9) …
(v10) Donoghue's initial **research**, published in the science journal Nature in 2002, consisted of attaching an implant to a monkey's **brain** that enabled it to play a simple pinball computer game remotely.
(v11) The four-millimeter square **chip**, which is placed on the surface of the motor cortex area of the **brain**, contains 100 electrodes each thinner than a hair which detect neural electrical activity. The sensor is then connected to a computer via a small wire attached to a pedestal mounted on the skull.
(v12) …
(v13)…
(v14)…
(v15) "Here we have a **research** participant who is capable of controlling his environment by thought alone -- something we have only found in science fiction so far," said Friehs.
(v16) …

- ***Page Graph (undirected, weighted)***
  - ❖ Text Fragments(nodes)
  - ❖ Semantic links(edges)
- ✓ Parsing delimiter – NewLine.
- ✓ Text Fragments – Paragraphs.
- ✓ 17 text fragments (v0…v16).
- ✓ 17 nodes in Document Graph.

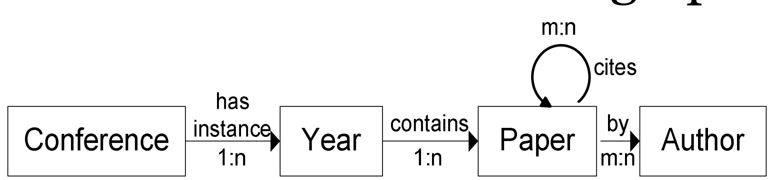


Data Model

- ***Data Graph* (directed, unweighted)**
  - ❖Tuples(nodes) & primary/foreign key relationships(edges).
  - ❖Each node represents an object & has a role.
  - ❖Each edge is labeled with its role.
  - ❖ Richer semantics – metadata.

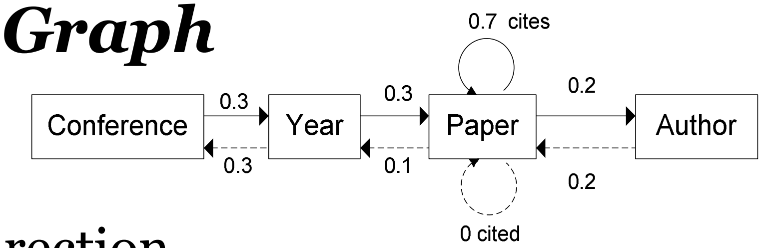


- ***Schema Graph***
  - ❖Describes the structure of the data graph.



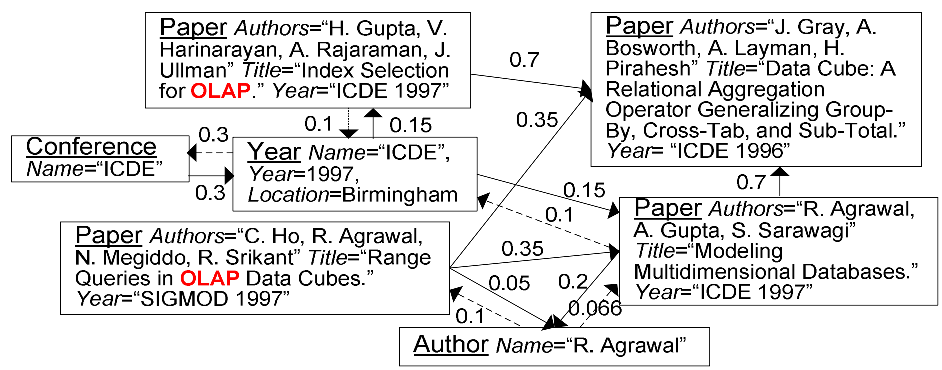
Data Model

- ***Authority Transfer Schema Graph***
  - ❖Edges reflect the authority transfer.
  - ❖Bi-directional authority transfer.
  - ❖Potentially different rates for each direction.



- ***Authority Transfer Data Graph* (directed, weighted)**
  - ❖Data graph edges labeled with authority transfer rates.

$$\alpha(e^f) = \begin{cases} \dfrac{\alpha(e_G^f)}{OutDeg(u,e_G^f)}, ifOutDeg(u,e_G^f) > 0 \\ 0, ifOutDeg(u,e_G^f) = 0 \end{cases}$$



Roadmap

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- **Related Work**
- Preliminary Work
- Ongoing & Future Work

Related Work

Overview

- Document Summarization.
- Keyword Search on Data Graphs.
- Traditional IR Ranking.
- Link-based Semantics.
- Relevance Feedback & Query Reformulation.

Related Work

1) Document Summarization

- ✓ Mostly Query-Independent
- ✓ Summarizing Web Pages
  - OCELOT - Berger et.al [SIGIR2000] synthesizes summaries (*non-extractive*).
  - INCOMMENSENSE - Paris et.al [CIKM2000] uses anchor text (ignores content).
- ✓ Splitting Web pages in to blocks
  - Song et.al [WWW2004] Block importance models (learning algorithms)
  - Cai et.al [SIGIR2004] Block level link analysis
- ✓ Document modeled as Graphs
  - Lexrank [JAIR2004] : Sentence Centrality using link analysis.
  - TextRank [EMNLP2004]: “representative” sentences using link analysis.

2) Keyword Search on Data Graphs

- BANKS [ICDE2002]: group-steiner tree problem
- DISCOVER[VLDB2002], DBXplorer[ICDE2002],IRStyle[VLDB2003].
- XRANK[SIGMOD2003], Xkeyword[ICDE2003]: search in XML documents.

Related Work

3) Traditonal IR Ranking

- Modern IR Overview.
  - Singhal [IEEE data bulletin 2001].

$$\sum_{t\in Q,d} \ln\frac{N-df+0.5}{df+0.5}\cdot\frac{(k_1+1)tf}{(k_1(1-b)+b\frac{dl}{avdl})+tf}\cdot\frac{(k_3+1)qtf}{k_3+qtf}$$

- Term weighting .
- State of art IR is based on tf *idf principle.
  - Okapi Formula.
  - Pivoted Normalized Weighting.

4) Link-Based Semantics

- PageRank [WWW98] for the Web.
- HITS [ACM Journal 99].
- Topic-Sensitive PageRank [WWW02] for the Web.
- ObjectRank for the database [VLDB02].
- XRANK [SIGMOD03] for XML databases.

Related Work

5) Relevance Feedback & Query Reformulation

- Salton, Buckley introduced Relevance feedback [Information Sciences 90].
- Term selection, re-weighting, query expansion [SIGIR94, TREC95].
- Ruthven, Lalmas - Complete Relevance feedback Survey [knowledge engineering review 2003]
- RF based on web-graph distance metrics [SIGIR06]
- Query-independent techniques to assign propagation factors - Nie et al. [WWW2005] , Agarwal et al. [SIGKDD2006]

Roadmap

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- **Preliminary Work**
- Ongoing & Future Work

Preliminary Work (published/accepted)

- **Structure-Based Query-Specific Document Summarization**
  - *Ramakrishna Varadarajan, Vagelis Hristidis*
  - Published in ACM CIKM, 2005 (2-page poster)
- **Searching the Web Using Composed Pages**
  - *Ramakrishna Varadarajan, Vagelis Hristidis, Tao Li*
  - Published in ACM SIGIR, 2006 (2-page poster)
- **A System for Query-Specific Document Summarization.**
  - *Ramakrishna Varadarajan, Vagelis Hristidis*
  - Published in ACM CIKM, 2006 (full paper)
- **Beyond Single-Page Web Search Results.**
  - *Ramakrishna Varadarajan, Vagelis Hristidis, Tao Li*
  - Accepted for publication in IEEE TKDE, 2008 (Journal paper)
- **Explaining and Reformulating Authority Flow Queries.**
  - *Ramakrishna Varadarajan, Vagelis Hristidis, Louiqa Raschid*
  - Accepted for publication in IEEE ICDE, 2008 (full paper)

Preliminary Work

Specific Research Goals

1. ***Improve Web Search Results*** [CIKM2005, SIGIR2006, CIKM2006, TKDE2008]
   - Improve Result Presentation.
   - Go beyond page-granularity.
   - Make it more user-friendly by reducing user-browsing time.
   - Improve the quality of results.

2. ***Improve Authority-Flow Based Graph Search*** [ICDE2008]
   - Make it more user-friendly.
   - Explain query results in an intuitive manner.
   - Personalize the search system.
   - Enable user-feedback.

Preliminary Work Overview

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
  - ❖ **Query-Specific Document Summarization**
  - ❖ Beyond Single-Page Web Search Results
  - ❖ Explaining & Reformulation Authority Flow Queries
- Ongoing & Future Work

Preliminary Work

Document Summarization **[CIKM'05,CIKM'06]**

- Locating relevant information is hard.
- **Summaries** are **helpful** because:
  - Provide a Quick preview of the document.
  - Allow users to quickly decide relevance.
  - Save user's browsing time.

- Success of *Web search engines* – Query specific **snippets** are important.
- Two categories of summaries:
  - *Query-Independent* – Most of prior works.
  - *Query-Specific* – Applicable to web search engines.

Preliminary Work

<u>Document Summarization</u> **[CIKM'05,CIKM'06]**

- Document → ***graph***
- We call it ***Document Graph***.

Three Steps

Step 1: ***Preprocess***

- Build a document graph, *G*.

Step 2: ***Summary Generation***(keyword proximity search)

- Given a query Q and a document graph G,

  Summaries → *Spanning Trees* that cover all keywords

Step 3: ***Rank*** spanning trees.

Preliminary Work

<u>Input parameters for *Document Graph* construction</u>

- ***Parsing*** Delimiters
  - For Plain Text – Newline or Period
  - For HTML – Tags (<p>,
,
,<table>… etc.)
- ***Threshold*** for Edge weights
  - Tradeoff of Quality and Performance.
  - Edges with weights lesser, are not added.
- ***Maximum*** Fragment Size
  - Limit on Node Size

<u>Edge Scoring</u>

$$EScore(e) = \frac{\sum_{w \in (t(u) \cap t(v))} \left((tf(t(u),w) + tf(t(v),w)) \cdot idf(w) \right)}{size(t(u)) + size(t(v))}$$

A ***tf*idf*** adaptation.

– *Query Independent*.

<u>Node Scoring:</u> *Query-dependent* (based on okapi formula).

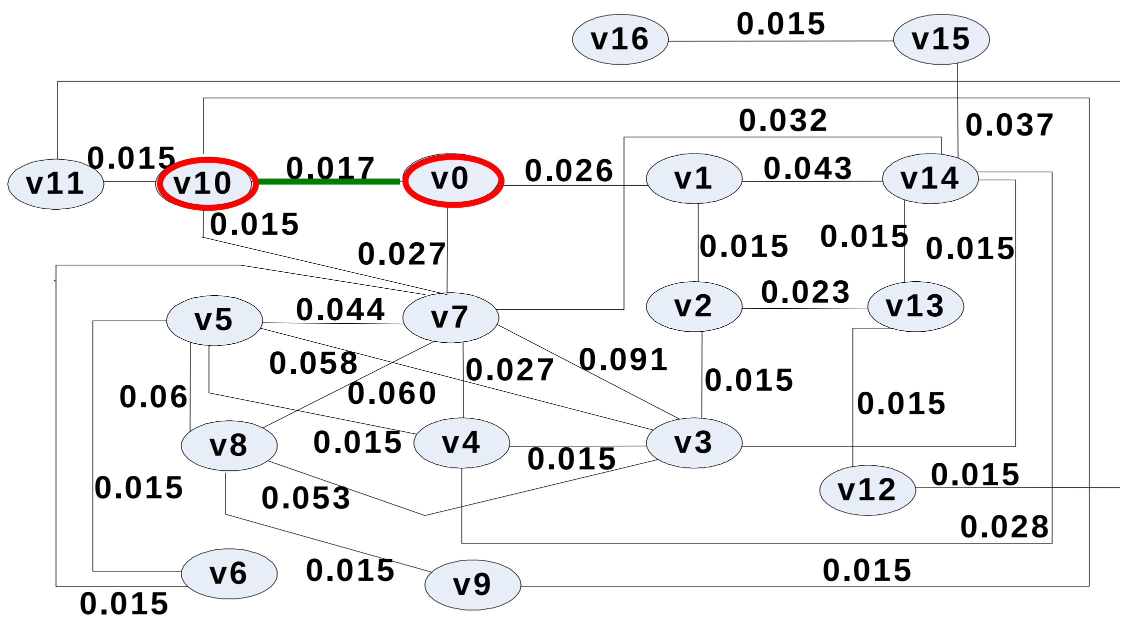
Preliminary Work

Example

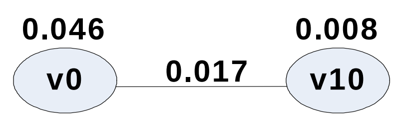
Sample Document

(v0) Brain chip offers hope for paralyzed
(v1) A team of neuroscientists have successfully implanted a **chip** into the **brain** of a quadriplegic man, allowing him to control a computer.
(v2) …
(v3) …
(v4) …
(v5) BrainGate offers the possibility of hitherto unimaginable levels of independence for the severely disabled.
(v6) …
(v7) …
(v8) …
(v9) …
(v10) Donoghue's initial **research**, published in the science journal Nature in 2002, consisted of attaching an implant to a monkey's **brain** that enabled it to play a simple pinball computer game remotely.
(v11) The four-millimeter square **chip**, which is placed on the surface of the motor cortex area of the **brain**, contains 100 electrodes each thinner than a hair which detect neural electrical activity. The sensor is then connected to a computer via a small wire attached to a pedestal mounted on the skull.
(v12) …
(v13)…
(v14)…
(v15) "Here we have a **research** participant who is capable of controlling his environment by thought alone -- something we have only found in science fiction so far," said Friehs.
(v16) …

Document Graph



Top Summary for "***Brain Chip Research***"



Score = **67.74**

Brain chip offers hope for paralyzed.
⌊ Donoghue's initial **research** published in the science journal Nature in 2002 consisted of attaching an implant to a monkey's **brain** that enabled it to play a simple pinball computer game remotely.

Preliminary Work

Summary Scoring Function Requirements

Properties of Good Summaries :

- Highly relevant nodes (fragments) ***improve*** Score.
- Loose semantic Links ***degrade*** Score.
- Large spanning trees get a ***degraded*** Score.
- Based on *Query-dependent* & *Query-Independent* factors.

Summary Scoring

- This function *satisfies* these requirements.
- Best Summary has ***minimum*** score

$$\text{Score(T)} = a \sum_{edge\ e \in T} \frac{1}{EScore(e)} + b \frac{1}{\sum_{node\ v \in T} NScore(v)}$$

a and ***b*** are calibrating parameters.

(a=1 & b=0.5)

Preliminary Work Overview

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
  - Query-Specific Document Summarization
  - **Beyond Single-Page Web Search Results**
  - Explaining & Reformulation Authority Flow Queries
- Ongoing & Future Work

Preliminary Work

Composed Pages Search [SIGIR'06,TKDE'08]

Motivation

- Current Web search engines return a list of ***individual*** web pages.
- Basic unit for search & retrieval - individual web page.
- Information – distributed across pages & are hyperlinked.
- Degrades quality of search results.
  - Especially for Long & Uncorrelated Queries.
- Li et al. [WWW01] ("Information Unit" paper).
- **We extract & stitch together pieces of information**
- **In contrast, we go beyond page granularity.**

Preliminary Work

Composed Pages Overview: [SIGIR'06,TKDE'08]

STEPS

1) Preprocessing : Web Page → labeled, weighted ***Page graph***.

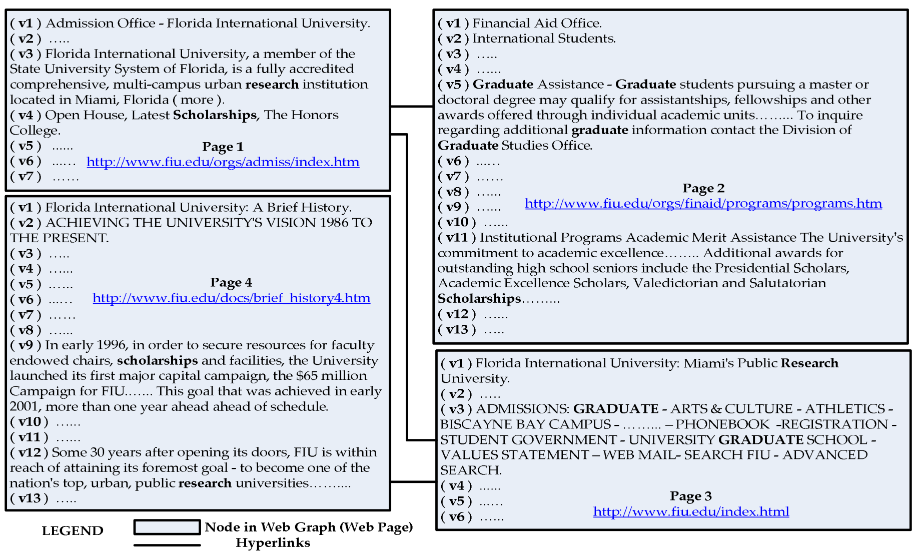
2) At Query Time: Given a set of keywords, Q.

- Compute *Web Spanning Tree* (**WST** – a hyperlinked set of pages).
- WST is total & minimal.
- Compute *Page Spanning Tree* (**PST** – a query-specific summary) for each page of WST.
- WST & PST computed using keyword proximity search.
- Appropriately combined → ***COMPOSED PAGE***
- Top-k Composed Pages are returned.

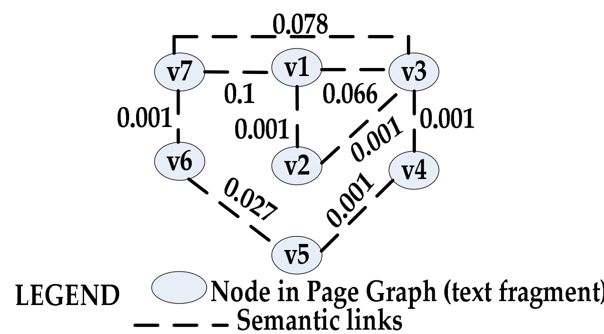
Preliminary Work

Composed Pages Example: [SIGIR'06,TKDE'08]

Web Graph (crawled)



Page Graph (pre-computed)



Rank	Score	Search Results
1	12.50	1: v3 -- v4 — 2: v5
2	101.60	3: v3 -- v1 — 4: v9
3	209.89	3: v3 -- v1 — 1: v4

Preliminary Work

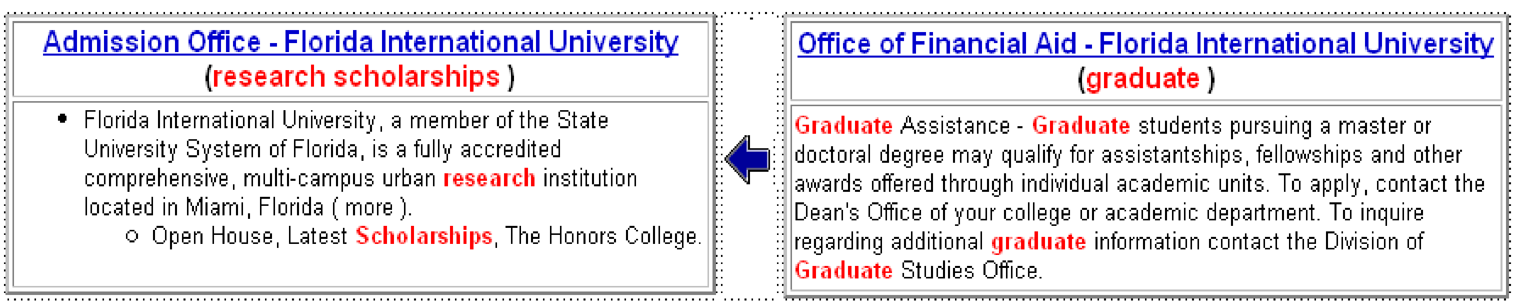
Presentation & **Ranking** of Composed Pages

- First ranking principle - search results involving fewer pages are ranked higher.

$$Score\ (R) = \sum_{p \in R} \frac{Score\ (p)}{PR\ (p)}$$

- Second ranking principle - Within search results of same page size, rank according to the involved page spanning trees.

- Scores of PSTs are combined using a monotone combining function.



Preliminary Work

<u>ALGORITHMS & EXPERIMENTS</u> **[SIGIR'06,TKDE'08]**

- Adaptations of **BANKS [ICDE02]** Algorithms
- ***Enumeration*** Algorithm.
- ***Expanding Search*** Algorithm.
- **Pre-computation**:
  - A Full text Index.
  - PageRank values of each web page.
  - Page Graphs of each web page.
  - All Pairs shortest paths for each page graph (edge weight of edge e= 1/Escore(e))
- User Surveys – DUC, Google/MSN Desktop.

Preliminary Work Overview

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
  - ❖Query-Specific Document Summarization
  - ❖Beyond Single-Page Web Search Results
  - ❖ **Explaining & Reformulation Authority Flow Queries**
- Ongoing & Future Work

Preliminary Work

Explaining & Reformulating Authority Flow Queries [ICDE'08]

Motivation:

Limitations of ObjectRank [VLDB04] :

- No way to *explain* to the user why a particular result received its current score.
- Authority transfer rates have to be set manually by a domain expert.
- No *query reformulation* methodology to refine results.

Focus

- Web search (out of scope) - we focus on typed domain -specific data graphs.

Preliminary Work

Query Definition & ObjectRank2 [ICDE'08]

- A keyword query Q is defined as a tuple of keywords $Q=[t_1,\ldots,t_m]$
- For each query $Q=[t_1,\ldots,t_m]$ we define a *query vector* $\boldsymbol{Q}=[w_1,\ldots,w_m]$
- Random Surfer jumps to different nodes of base set with different probabilities.
- Probability for a node v is proportional to I*RScore*(v,$\boldsymbol{Q}$)

$$r^Q = dAr^Q + \frac{(1-d)}{|S(Q)|}s$$

Power Method

A (transition matrix),
S (base set), s (base set vector),
r (objectrank2 scores vector)

Preliminary Work

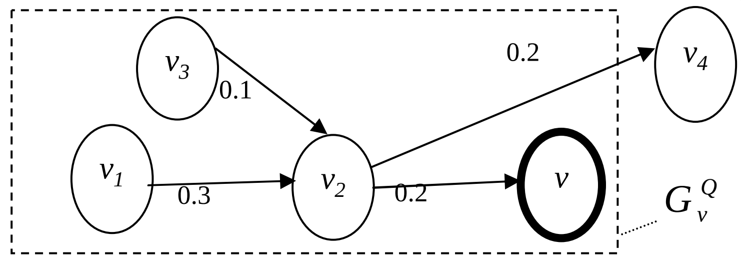
Explaining Authority-Flow Query Results [ICDE'08]

- Problem – Given a *target object* T, explain user why it received a high rank (or score).
- Our Solution – Display an explaining subgraph of Authority transfer data graph, for T.
- Explaining subgraph contains:
  - All Edges that transfer authority to T.
  - Edges are annotated with amount of authority flow.
- Steps:
  - Construction Stage (construct using Breadth-First Search from Base set S)
  - Flow Adjustment Stage (adjust original authority flows – most challenging)

Preliminary Work

- **Flow Adjustment Stage**

  1) Intuition



  2) Original Authority Flow $Flow_0(v_i \rightarrow v_j) = d \cdot \alpha(v_i \rightarrow v_j) \cdot r^Q(v_i)$

  3) Reduced Authority Flow $Flow(v_j \rightarrow v_k) = h(v_k) \cdot Flow_0(v_j \rightarrow v_k)$

  4) Reduction Factor $\mathrm{h}(v_k) = \dfrac{r^{Q\,\prime}(v_k)}{r^Q(v_k)}$

Important Find:

$$\mathrm{h}(v_k) = \sum_{(v_k, v_j) \in G_v^Q} \left(h(v_j) \cdot \alpha(v_k \rightarrow v_j) \right)$$

- *The "original" ObjectRank2 scores are **NOT** used in computing the reduction factor h(v_k).*

Preliminary Work

Query Reformulation

- Well studied in Traditional IR (Salton, Buckley 1990)
- Query Expansion was the dominant strategy (ignores link-structure)

STEPS:

1) System computes Top-k objects with high ObjectRank2 scores.
2) User marks relevant "feedback" objects.
3) Compute explaining subgraph of feedback objects.
4) Reformulate based on (a) **Content** (b) **Structure**.
5) Practically diameter is limited to a constant (L=3).

Roadmap

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
- **Ongoing & Future Work**

Preliminary Work Overview

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
- Ongoing & Future Work
  - **Information Discovery in Clinical Databases**
  - XML Search Evaluation
  - Flexible & Efficient Querying on hyperlinked Data

Ongoing & Future Work

Specific Research Goals

Information Discovery on Clinical Databases

- Objective - Develop methods to effectively search EMRs.
- Design - Clinical ObjectRank (CO) System.
- Consider domain-specifics for ranking.
- Personalize the system for a variety of users.

IMPLEMENTATION STEPS

- *DATASET* - EMR dataset of cardiovascular division of MCH.
- Measurements – Precision/recall comparing traditional IR.
- Customization – Develop user profiles for a researcher, physician, pharmacist, nurse, therapist,……

Preliminary Work Overview

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
- Ongoing & Future Work
  - Information Discovery in Clinical Databases
  - **XML Search Evaluation**
  - Flexible & Efficient Querying on hyperlinked Data

Ongoing & Future Work

Specific Research Goals

XML Search Evaluation

- Objective - Create distance measures for Top-k XML lists.
- Method – based on tree-edit distances.
- XML Top-k Distance Algorithms based on
  - (a) Spearman's footrule
  - (b) Kentall Tau
- Formally prove the distance metric conditions.
- XML List Aggregation based on the proposed distance metrics
- Finally provide a **Case Study**.

Preliminary Work Overview

- Problem Statement & Motivation
- State of Art Graph Search Methods
- Data Model
- Related Work
- Preliminary Work
- Ongoing & Future Work
  - Information Discovery in Clinical Databases
  - XML Search Evaluation
  - **Flexible & Efficient Querying on hyperlinked Data Sources**

Ongoing & Future Work

Specific Research Goals

Flexible & Efficient Discovery over hyperlinked data [SIGMOD 2008 submission]

- Create a simple & intuitive framework.
- Make it flexible & extensible.
- Optimize the search execution.
- Support a variety of users – from sophisticated to naïve.

Joint work with 4 collaborators:

1. Framework/Query language based on Soft & Hard Filters.
2. Closed algebra of physical operators / rewriting rules.
3. Exact / approximate optimizations techniques for authority flow based soft filters.

Thank You !!!

Questions ???