

Fusion of Heterogeneous Information in Graph-Based Ranking for Query-Biased Summarization

Kotaro Sakamoto[†], Hideyuki Shibuki[†], Tatsunori Mori[†], Noriko Kando[‡]
[†]Yokohama National University, [‡]National Institute of Informatics

Background and Purpose

Accessing user-preferred information → Query-Biased Summarization
 Sentence extraction-based summarization → Graph-Based Ranking
 1. **Sentence importance** → Graph-Based Ranking
 2. Sentence selection
 Conventional method:
 The link weight based on the degree of **word-overlap**

Proposed Method

Precision Improvement

three-layer graph model based on **Basic Element overlap similarity**, **Semantic Similarity**, **Part-whole relationships**
 (Co-HITS-Ranking: two-layer graph model based on **word-overlap**)

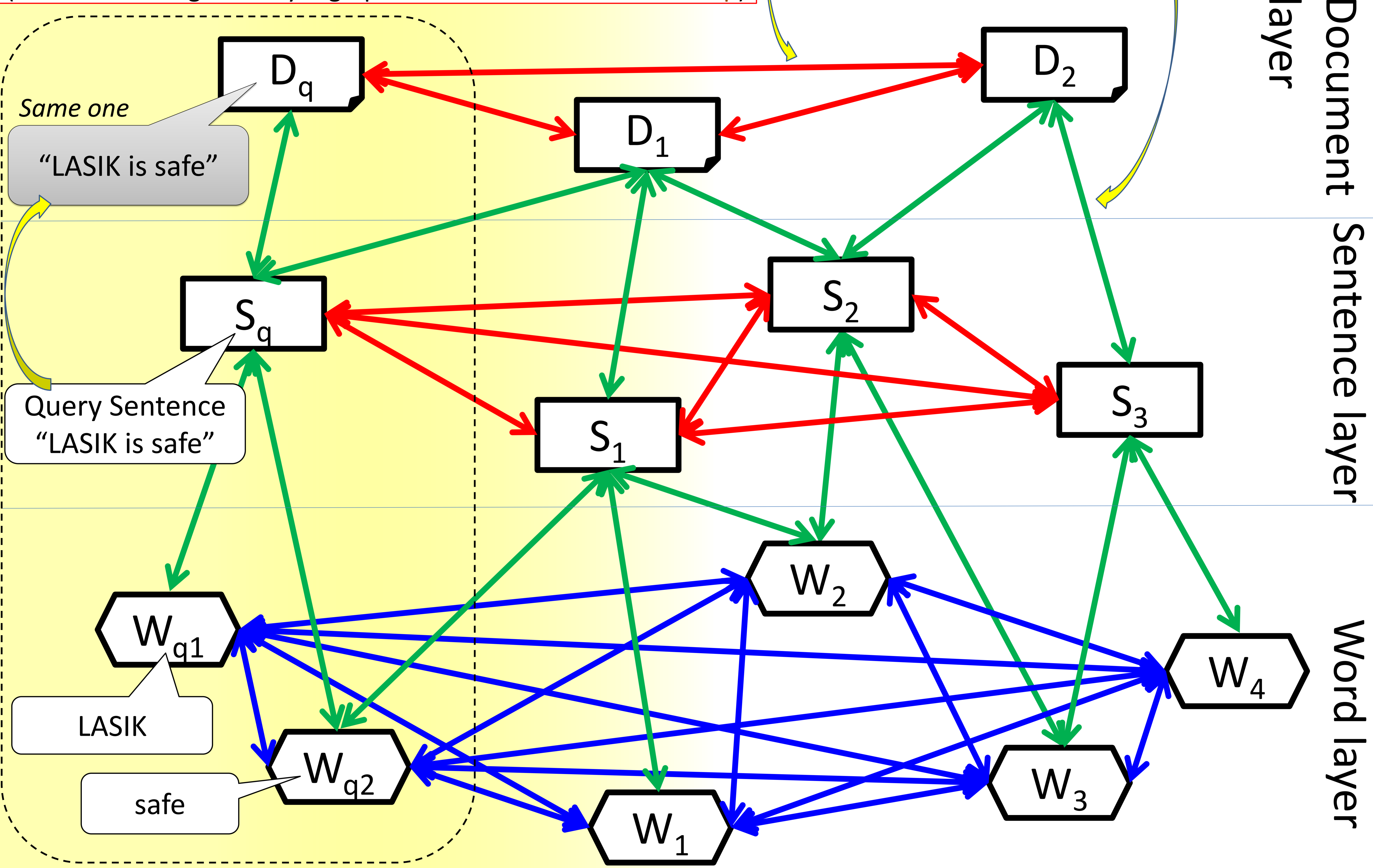
Co-HITS-Ranking Algorithm (Hu et al. 2010)

i. If n is a query node $R^q(n)=1$, otherwise 0
 ii. Among homogeneous nodes, \leftrightarrow

$$R^o(n_i) = d^o \sum_{n_j \in \text{In}(n_i)} \frac{w_{ij}^o}{\sum_{n_k \in \text{Out}(n_j)} w_{jk}^o} R^o(n_j) + (1-d^o)R^q(n_i)$$

 iii. Among heterogeneous nodes, \updownarrow

$$R^e(n_i) = d^e \sum_{n_j \in \text{In}(n_i)} \frac{w_{ij}^e}{\sum_{n_k \in \text{Out}(n_j)} w_{jk}^e} R^e(n_j) + (1-d^e)R^o(n_i)$$



Fusion of Heterogeneous Information

BoW-like effect

Basic Element overlap similarity

$$\text{sim}_{\text{BE}}(n_1, n_2) = \frac{|\text{BE}(n_1) \cap \text{BE}(n_2)|}{|\text{BE}(n_1) \cup \text{BE}(n_2)|}$$

Semantics Similarity

$$\text{sim}_{\text{sem}}(n_1, n_2) = \frac{\text{MD} - \min_{c \in \text{hyp}(c_1, c_2)} (\text{dep}(c))}{\text{MD}}$$

Part-whole relationships
 If upper unit contains lower unit 1
 otherwise 0

Basic Element
 Minimal semantic unit (Hovy et al. 2005)
 Dependency Structure Analyzer CaboCha results:
 watashi wa / eigo wo / benkyo suru
 (I) (English) (study) = I study English
 (benkyo | watashi | agent)
 (benkyo | eigo | object)

Experiment

Three layers vs Two layers vs One layer
 Text summarization test collection of Nakano et al. (Nakano et al. 2010)
 6 statements,
 4 summarizers / statement
 250 Web documents / statement

Average R Precision of 6 statements ($d^o = d^e = 0.5$)

Only S-layer	0.173	≡ Zhou, 2005 Not graph-based
S+D-layer	0.151	≡ Co-HITS-Ranking
S+W-layer	0.240	Part-whole + BE
Three layers	0.338	proposed