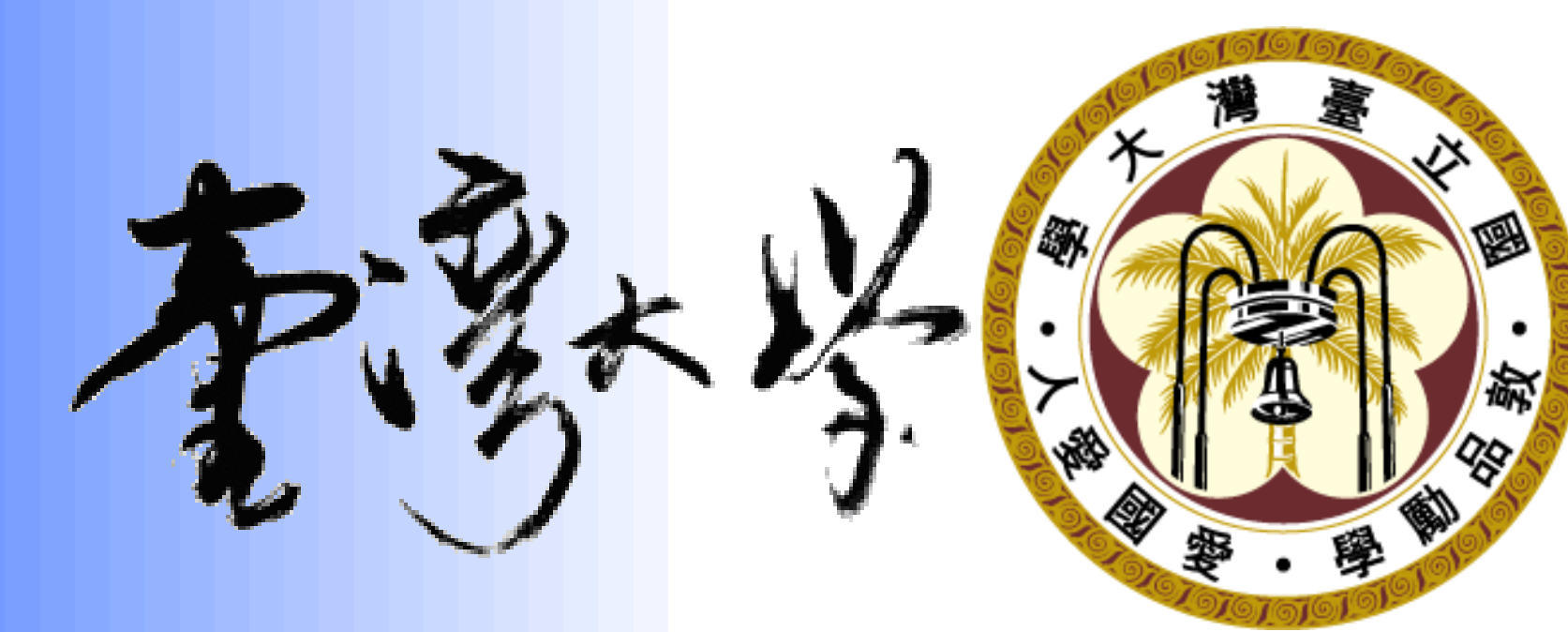


Spoken Question Answering using Tree-structured Conditional Random Fields and Two-layer Random Walk

Sz-Rung Shiang, Hung-yi Lee, Lin-shan Lee



National Taiwan University

Introduction

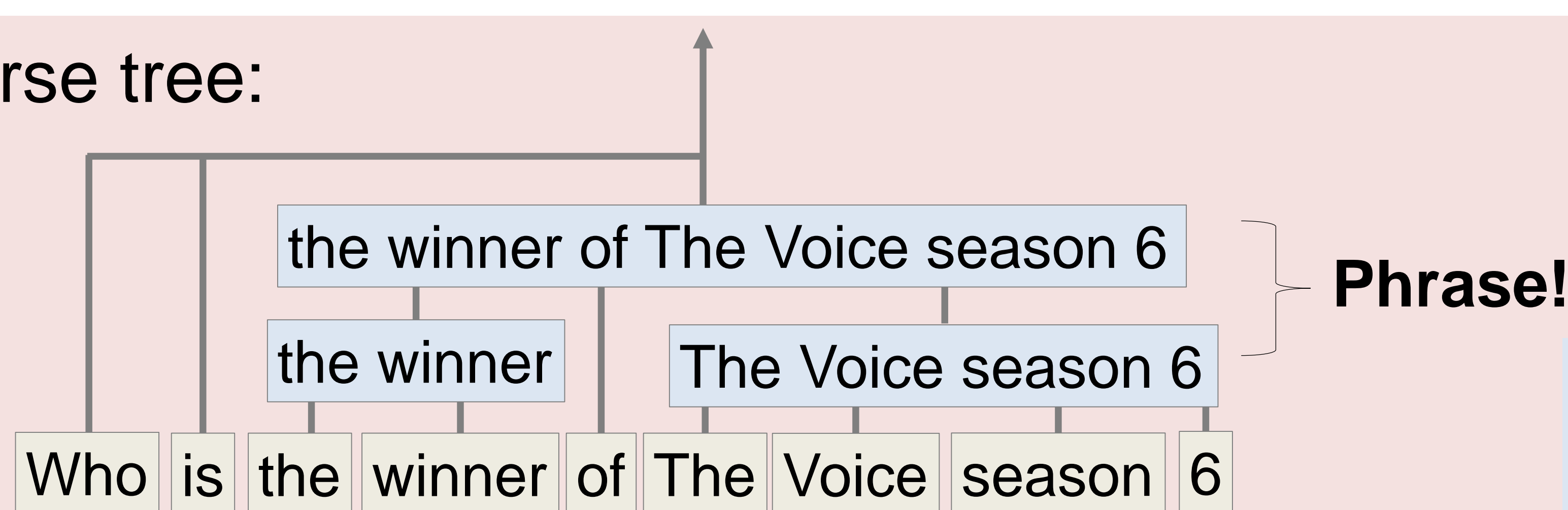
➤ Question Answering: spoken queries / text answers from retrieved text documents

➤ Utilize **phrases** in **parse tree** and leverage parse tree structure in **Conditional Random Fields (CRF)** to form the queries for the questions.

➤ Example question:

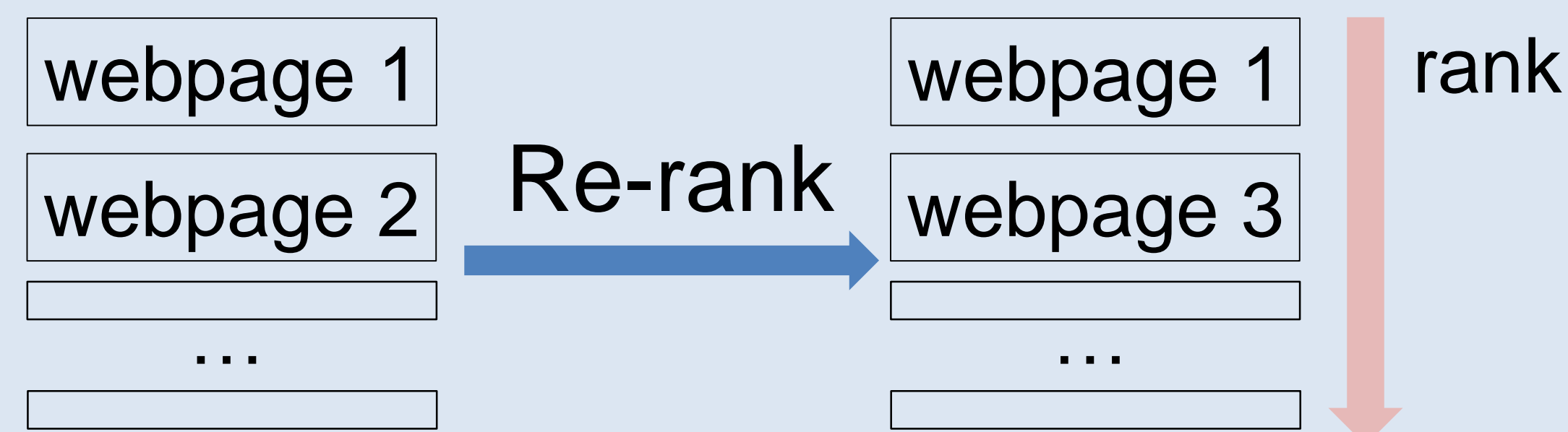
“Who is the winner of The Voice season 6?”

➤ Parse tree:

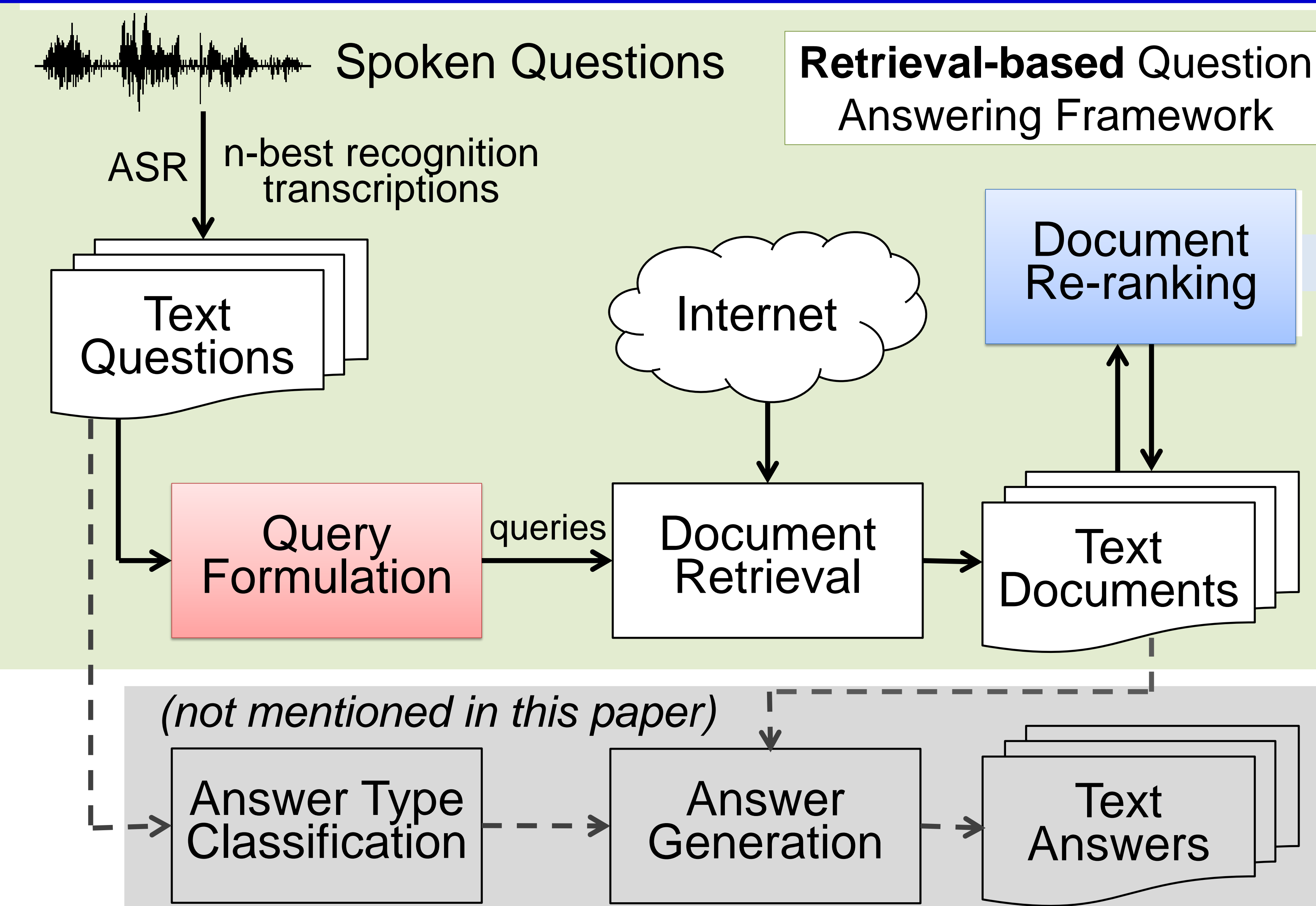


➤ Re-ranking the webpages retrieved by queries from n-best using **Two-layer Random Walk**.

➤ Webpages containing correct answers ranked higher.

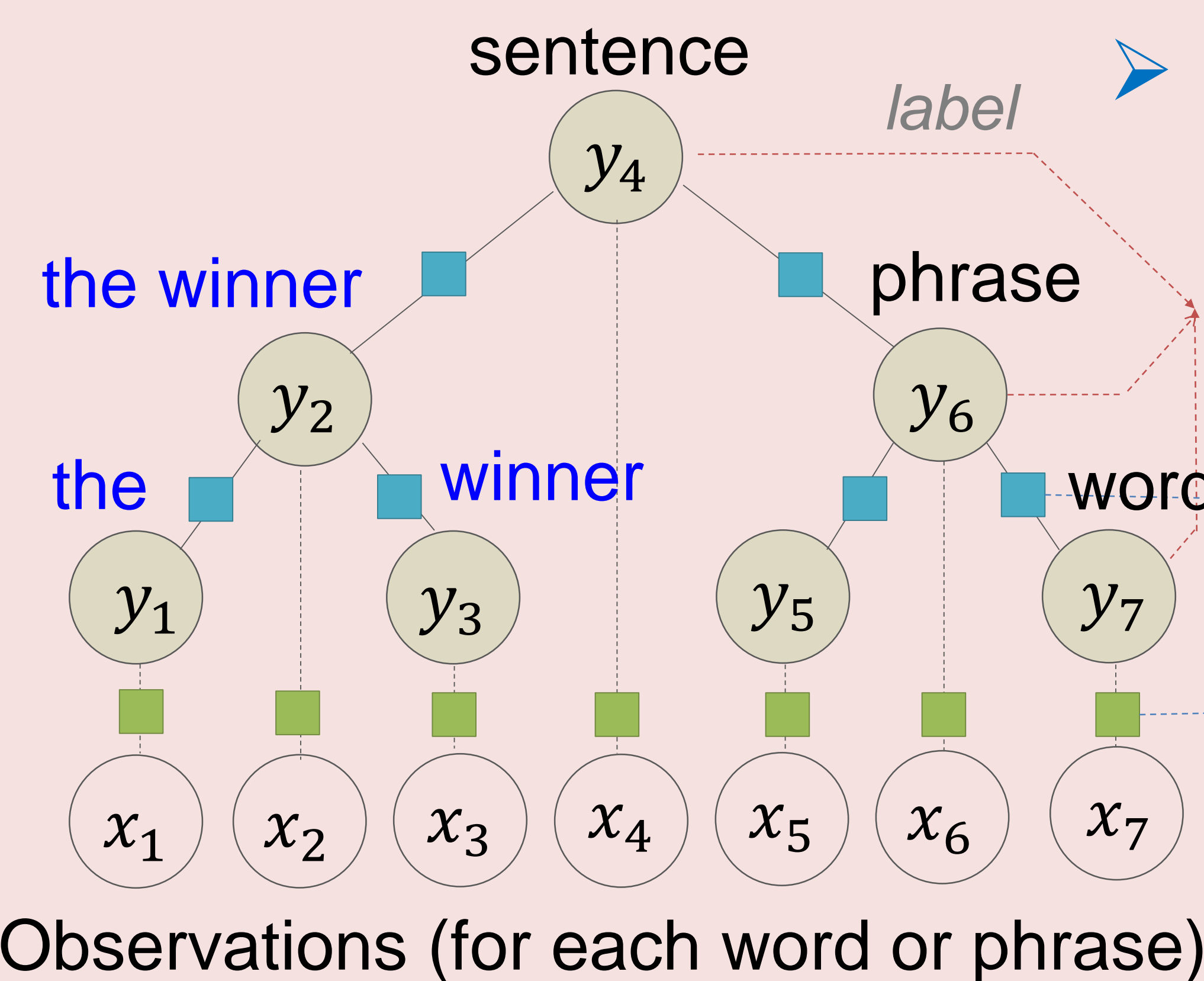


Framework



Proposed Approach

1. Query Formulation Leverage parse tree structure in Conditional Random Fields (CRF).



➤ Formulate the queries from spoken questions (n-best transcriptions)

➤ Parse tree structure: each node in CRF is a word or a phrase.

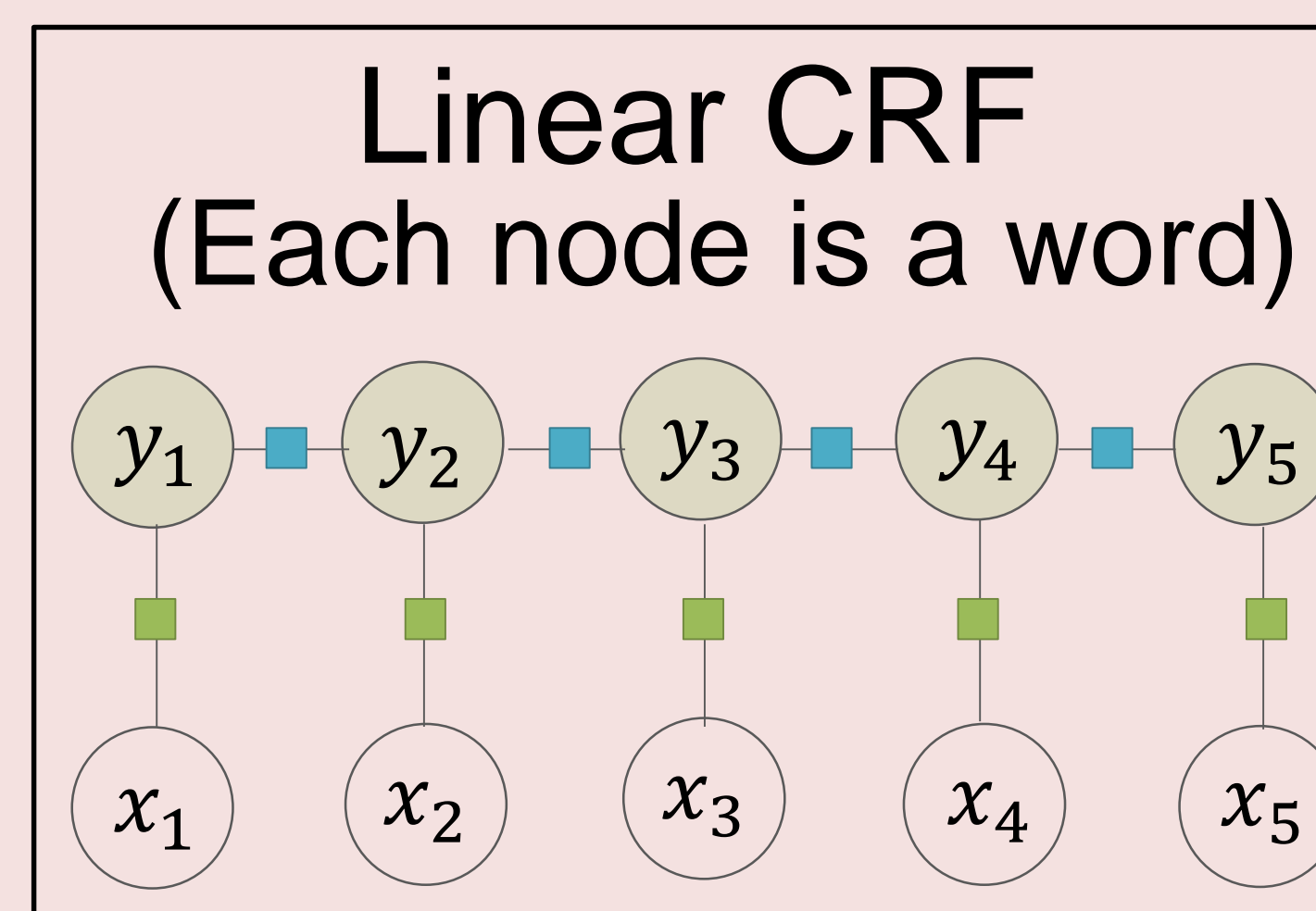
➤ Desired labeling: 1. answer related (for further use)

2. query related

3. none of the above

➤ Tree structured CRF and feature functions: details are in paper.

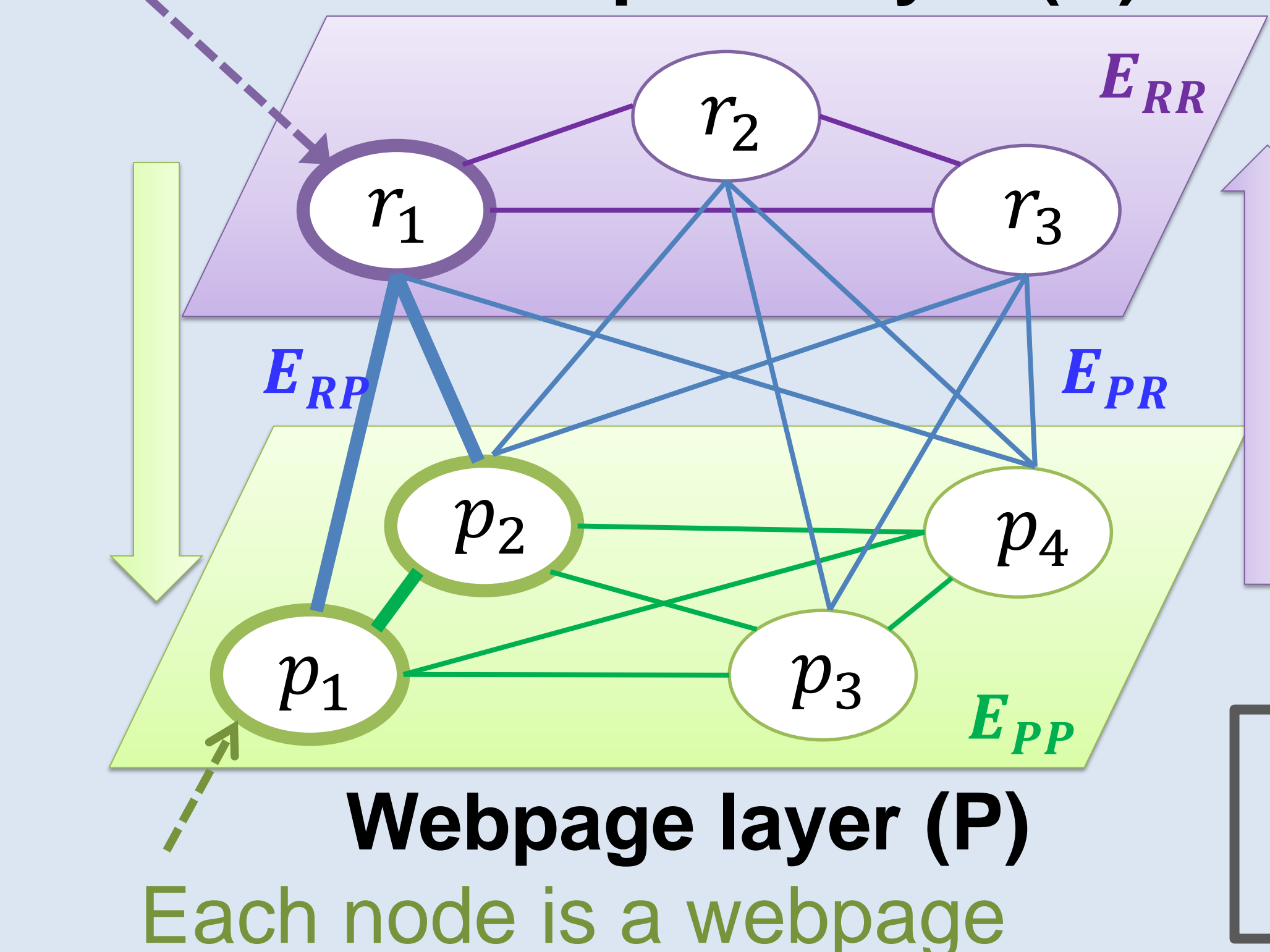
➤ Query formulation: cascading phrases and words labeled as “query related”



2. Document Re-ranking

➤ **Two-Layer Random Walk:** Scores propagate over the two-layer graph mutually enhanced by similarity

Each node is one of n-best transcripts transcription layer (R)



➤ Iteratively update score of layers:

E_{PP} : (similarity within webpage layer) cosine similarity between webpages

$F_R^{(t)}$: transcription scores at t -th iteration= $[r_1, r_2, r_3]$

$$F_P^{(t+1)} = (1 - \alpha) \cdot F_P^{(0)} + \alpha \cdot E_{PP}^T E_{PP} F_R^{(t)}$$

E_{RP}/E_{PR} : (similarity between layers) 1 if webpage is from the n-best transcript

$$F_R^{(t+1)} = (1 - \alpha) \cdot F_R^{(0)} + \alpha \cdot E_{RR}^T E_{RP} F_P^{(t)}$$

E_{RR} : (similarity within transcription layer) cosine similarity between queries from n-best.

$F_P^{(t)}$: webpage scores at t -th iteration= $[p_1, p_2, p_3, p_4]$

- First propagate score from webpage layer
- Then propagate score within transcription layer

➤ Convergence: $F_P^{(T+1)} = F_R^{(T)}$, $F_R^{(T+1)} = F_P^{(T)}$

Corpus:

- 189 question-answer pairs in Mandarin Chinese from quiz show.
- Recorded by a single speaker with total length about 58 minutes.
- 126 for training, 63 for testing. (3-fold)
- 12.19% WER and 59.26% SER.
- 5-best transcriptions used.
- **Precision (P) & Mean Average Precision (MAP)** as evaluation measures.

Experiments

	▼ Query Formulation		▼ Re-ranking Webpages			
	Linear CRF	Tree CRF	One-best	5-best	Two-layer random walk	
P@3	0.4746	0.5138	P@3	0.5021	0.5138	0.5183
P@5	0.4558	0.4793	P@5	0.4770	0.4793	0.5013
MAP@3	0.5716	0.6053	MAP@3	0.6099	0.6053	0.6239
MAP@5	0.5656	0.5968	MAP@5	0.5956	0.5968	0.6103