Optimizing Web Search for Learning
Incorporating Keyword Density as a Retrieval Objective

Rohail Syed
Kevyn Collins-Thompson

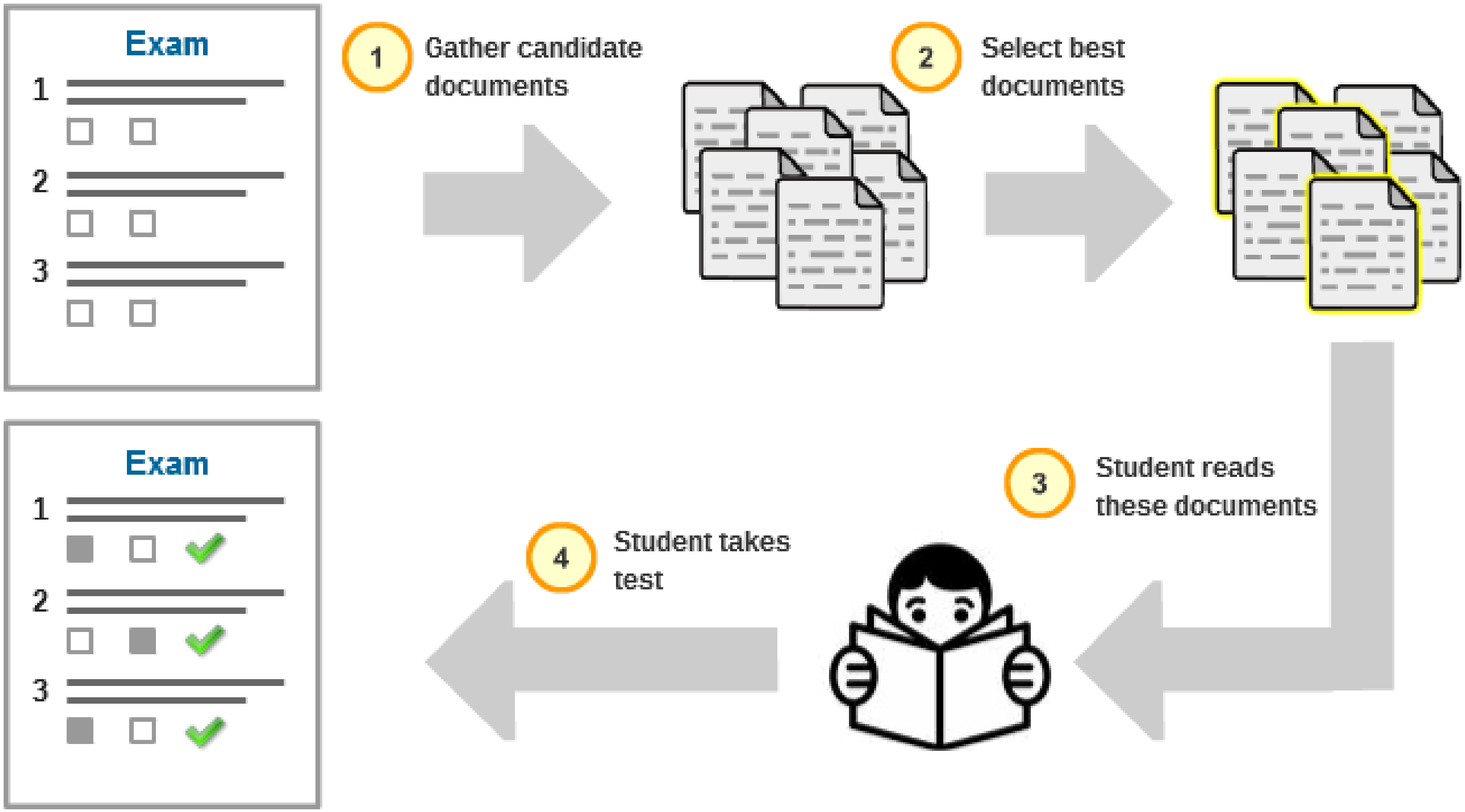
Background

- Web search commonly used as starting point for learning tasks
- Current ranking algorithms not designed to optimize learning utility
- Little existing work on Intelligent Tutoring Systems (ITS) applied to general web search

General ITS Framework

- **Student** wants to learn subject X
- **Expert** determines set of questions to test knowledge of X
- **Tutor** determines resources for student to perform best on test
  - Our focus. What web search algorithm optimizes the student's future test scores.

ITS Framework



Icon credit: *http://thumb10.shutterstock.com/photos/thumb_large/255466/111580295.jpg*

Expert Model

- **Expert** selects set of documents that are most representative of the topic that needs to be taught.
- In this study, Expert chose documents manually. Future work will use an automated selection procedure.

Tutor Model

- **Tutor's** objective is to teach the student a set of *N* topic-relevant vocabulary terms (keywords)
- These are determined using a weighted term-frequency measure on the Expert documents

$$Score_{Tki} = \frac{TF_i}{\log Total_i}$$

Tutor Model (Cont'd)

- Construct multinomial of weights W based on keyword frequency in Expert corpus.
- Our model assumes students are Bayesian learners.
  - Need to determine how to distribute T total keywords amongst the N keywords. Ideally $T = \infty$
  - So, we assign the student to read $S_i = W_i * T$ instances of the i^{th} keyword.

Objective Function

- Intrinsic Diversity objective[1]:

$$\arg\max_{\mathcal{D}} \sum_{i=1}^{|\mathcal{D}|} Rel(d_i|q) \cdot Rel(d_i|q_i) \cdot e^{\beta Div(q_i,\mathcal{Q})}$$

- ID objective + keyword density penalty

$$\arg\max_{\mathcal{D}} \sum_{i=1}^{|\mathcal{D}|} Rel(d_i|q) \cdot Rel(d_i|q_i) \cdot e^{\beta Div(q_i,\mathcal{Q})} \boxed{\cdot\, e^{\alpha \epsilon_i}}$$

Keyword density parameter

$$\epsilon_i = \frac{1}{|d_i|} \sum_{j=1}^{N} \begin{cases} C_{Dj} & C_{Dj} + C_{Rj} \leq C_{Tj} \\ \max(0, C_{Tj} - C_{Rj}) & \text{else} \end{cases}$$

- C_{Dj} = instances of j^{th} keyword in *D*
- C_{Rj} = total instances of j^{th} keyword seen so far
- C_{Tj} = total instances of j^{th} keyword required to be seen

Study Design

- Crowdsourced study (using CrowdFlower)
  - Five topics tested with four levels of α [0,80,120, ∞]. Each condition assigned 35 participants
  - Participants: (1) complete pre-reading test (2) read provided documents (3) complete post-reading test
  - Learning gain = sum of improvement in scores for each keyword

Results

- Mean learning gains per condition and topic

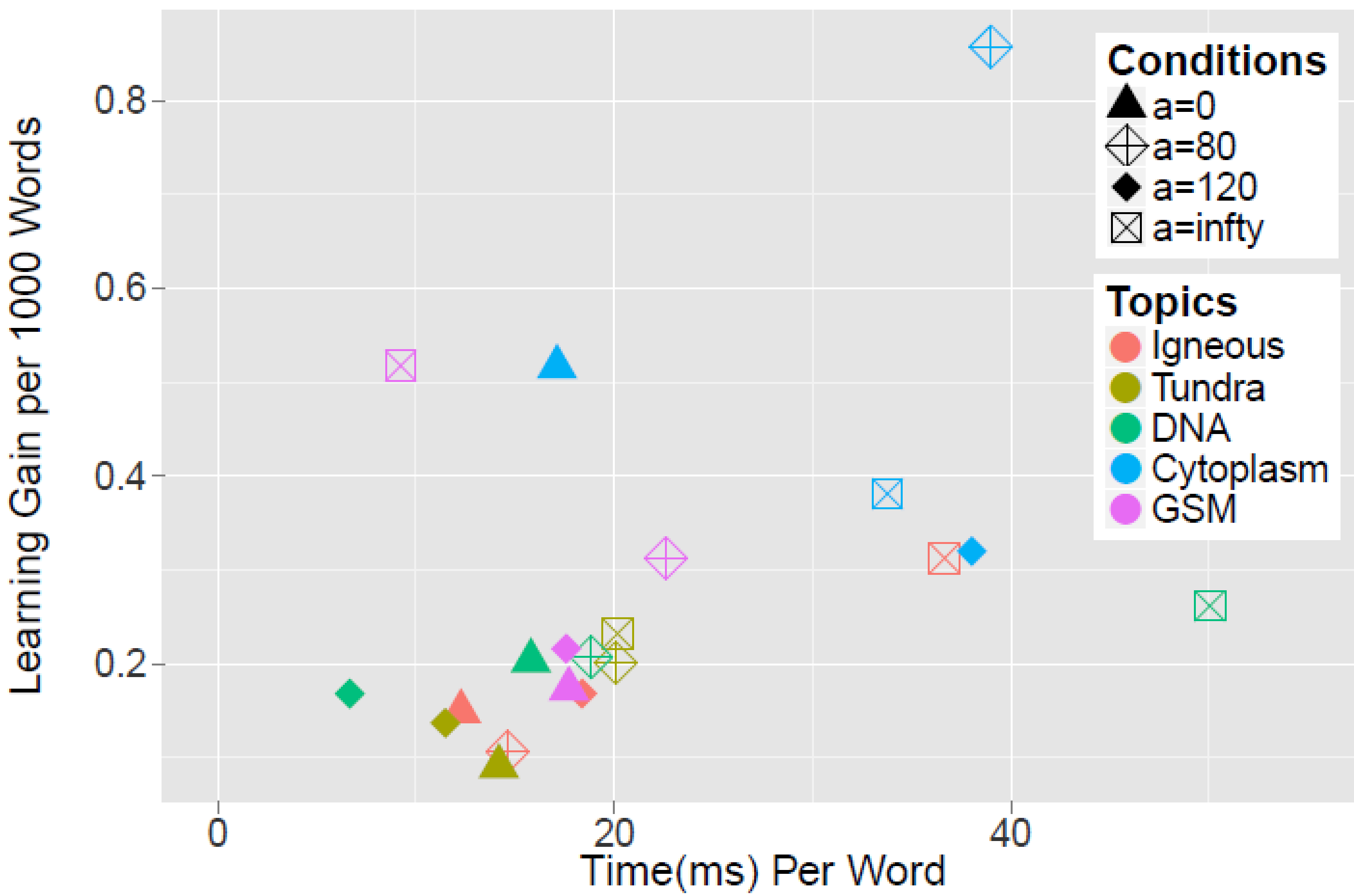
Topics	α = 0	α = 80	α = 120	α = infty	p-value
Igneous rock	1.312	1.094	1.333	**1.529**	p=.562
Tundra	1.406	**1.829**	1.800	1.514	p=.346
DNA	1.481	**1.576**	1.438	1.483	p=.977
Cytoplasm	1.719	**3.067**	1.286	1.333	p<.001***
GSM	1.654	**2.478**	1.258	1.967	p=.0126*
Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1					

Results

- Mean learning gains per 1000 words read per condition and topic

Topics	α = 0	α = 80	α = 120	α = infty	p-value
Igneous rock	0.149	0.106	0.168	**0.312**	p<.001***
Tundra	0.091	0.201	0.137	**0.232**	p<.001***
DNA	0.203	0.207	0.168	**0.261**	p=.258
Cytoplasm	0.516	**0.857**	0.320	0.381	p<.001***
GSM	0.173	0.312	0.216	**0.517**	p<.001***
Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1					

Learning Per Word vs Time Per Word



Conclusion

- We construct objective function in ITS setting to optimize learning gains.
- We find early evidence that keyword density objective can improve actual learning.
- We consider importance of incorporating other media types in objective function.

References

1. Raman K., Bennett P. N., and Collins-Thompson K. 2013. Toward Whole-session Relevance: Exploring Intrinsic Diversity in Web Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. ACM, New York, NY, USA, 463-472.
2. De Angeli A., Coventry L., Johnson G., and Renaud K. 2005. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies* 63, 1 (2005), 128-152.

Other Results

- Consider possible interaction of pictures in learning process
  - Picture superiority effect[2]. We were testing "Remember" learning (vocabulary learning)
  - Higher α conditions had higher image coverage