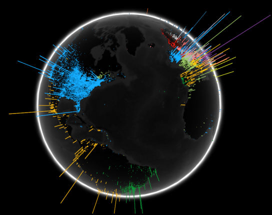
Contextual Advertising

Hari Sundaram
Associate Professor (CS, ADV)
hs1@illinois.edu

thanks: Andrei Broder, Vanja Josifovski

Introduction

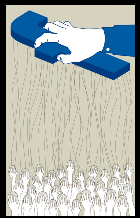


Web search

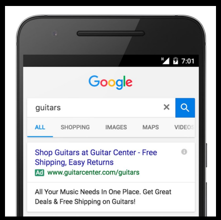
Game Theory

Auctions

Data flows



Privacy



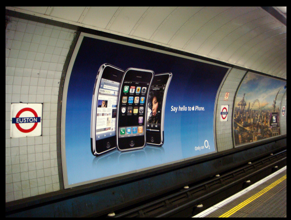
Text Ads

Contextual & Display Ads

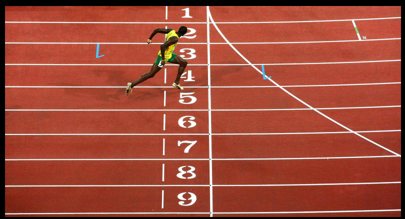
NETFLIX

Recommender systems

Behavioral targeting

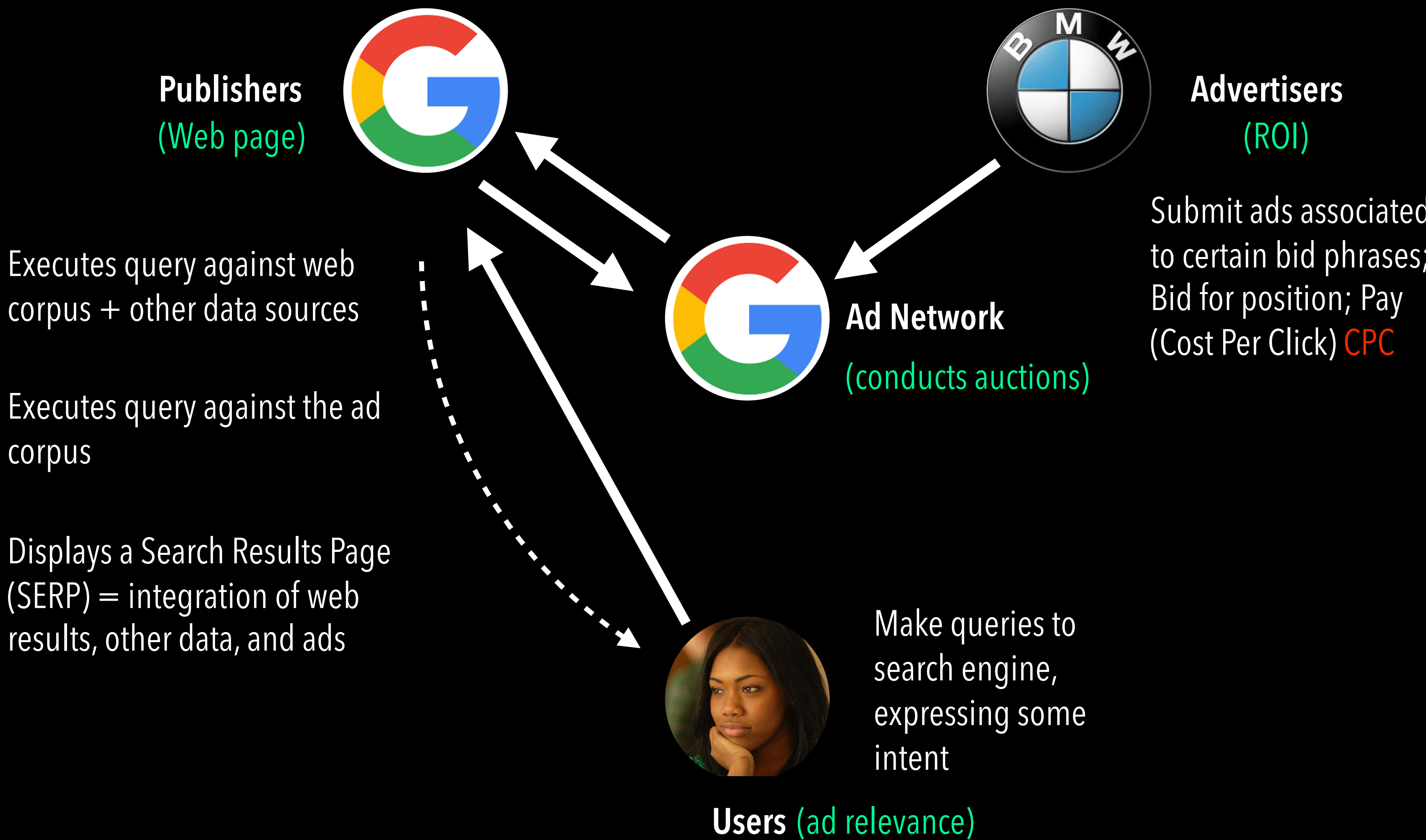


Emerging areas



Final Presentations

Sponsored Search



Textual advertising on third party web pages

- Complement the content of the web page with paid content
- Ubiquitous on the web

Main goal is to increase volume for textual campaigns in sponsored search

Same type of ads (companion campaign for sponsored search)

Advertiser opts into contextual advertising

what are contextual ads?

Sites small and big rely on Content Match revenue to cover for the cost of existence

Today, many sites are transitioning to using affiliate links instead of ads

Coverage at 100% usually—do not want to leave empty slots on the page

- Trade-off with display advertising

differences with sponsored search

Lesser role of the ad network, increased role of the publisher

- **Ad Network:** which ads
- **Publisher:** how many/where/how

Ad selection using the content of a web page

- Much more text
- Less focused
- Less intentional

Very thin margin business

CTR very low—orders of magnitude, ranges in 0.001-0.1%.

Higher CTR variance

challenges

Lower conversions – less of a clear intent

High volume—many page views per day

More difficult ad placement –intention of the person visiting page is less clear

Lower earnings for the search engine:

1) lower bids
2) share revenue with the publisher

Other benefits: User tracking

content match ad selection

Page Content

Process the content of the page

Cannot be done on-line: crawl

Most flexible from the ad selection perspective

what content can a publisher provide?

Page Snippet

Part of the page

How much can we process online?

How much is enough?

Custom Keywords

Sponsored Search–like mechanism

Least flexibility in ad selection

More control for the publisher

Phrase extraction (from the publisher page)

- Map CM to Sponsored Search
- Extract phrases from the page
- Use these phrases to select ads (exact match or advanced match in Sponsored Search)
- Ads selected on a single feature (phrase) from the page and the ad
- Historically first approach

Two Strategies

IR approach

- Treat CM as a document similarity problem
- Pages are compared to the ads in corpus in a common feature space
- Bid phrase is one of the features used in matching
- Ads selected based on multiple (overlapping) features of the page and the ads

IR: a 10,000 foot view

Collection: Fixed set of documents

Query: Description of the user's information need

Goal: Retrieve documents with information that is relevant to user's information need and helps him complete a task

Sim(a,p) is a function of a set of features of a and p ; a and p are vectors

Usually calculate similarity in a high dimensional space of features. Orders of magnitude numbers for textual ads:

- unique words ~ 1M-2M
- sequences (phrases) ~ 10M

Similarity search

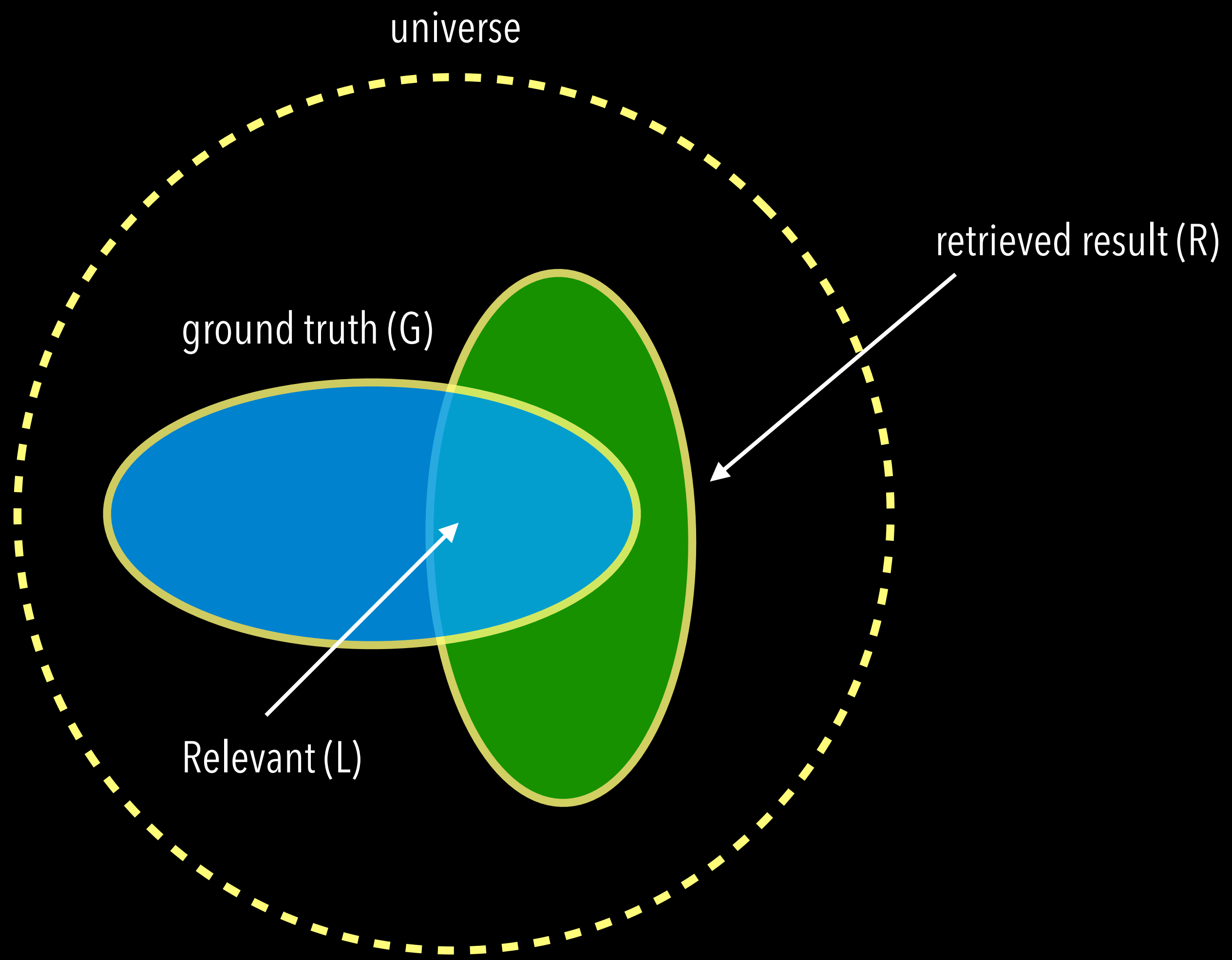
weight of the k-th
term in the ad

weight of the k-th
term in the query

dot product

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}||\vec{d}|} = \frac{\sum_k q_k d_k}{\sqrt{\sum_j q_j^2}\sqrt{\sum_j d_j^2}}$$

cosine of the angle
between the query
vector and the
document vector



metrics: precision / recall

$$P = \frac{|L|}{|R|} \qquad R = \frac{|L|}{|G|}$$

Phrase extraction for Contextual Advertising

Wen-tau Yih, Joshua Goodman, and Vitor R. Carvalho. 2006. Finding advertising keywords on web pages. In Proceedings of the 15th international conference on World Wide Web (WWW '06). ACM, New York, NY, USA, 213-222.

Reverse search problem:

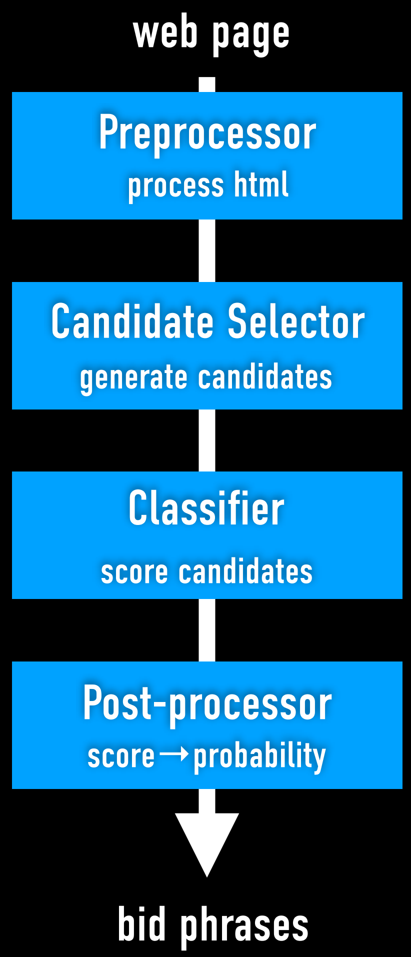
- Given a page, find the queries that would match (summarize) the content of this page

Goal: given a page, find phrases good for placing ads

Select ads based on a single selected keyword:

- Contextual Advertising translated into database approach of Sponsored Search
- Reuse of the Sponsored Search infrastructure—lower cost

system architecture



All phrases of length up to 5 (including single words)

- Within a single page block (sentence)

candidate selection

Two dimensions of candidate selection:

- Individual occurrences extracted separately vs. combining all occurrences into a single entry per page
- Consider the phrase as a whole
- Label individual words with their relationship with a phrase:
  - Beginning of a phrase
  - Inside a phrase
  - Last word of a phrase
  - …

Binary classifier

$$P(Y = 1 \mid X = x) = \frac{1}{1 + \exp(-wx)}$$

Logistic Regression

vector of importance weights

vector of features of a phrase

$$P(Y = 1 \mid X = x) = \frac{1}{1 + \exp(-wx)}$$

vector of features of a phrase

Linguistic features: is a noun; is a proper name; is a noun phrase; are all words in the phrase of the same type

Capitalization: any/all/first word capitalization

features

Section based features:

- Hypertext–is the feature extracted from
  - anchor text
  - Title
  - Meta tags
  - URL

IR features: tf, idf, log(tf), log(idf), sentence length, phrase length, relative location in the document

Query log features: log(phrase frequency), log(first/second/interior word frequency)

828 pages
Indexed by MSN
Have ads

data

In the Internet Archive (to allow for replicability)
One page per domain (for copyright and for diversity)
Eliminate foreign and adult pages
Editors (8) instructed to seek highly prominent keywords with advertising potential

Committee top-1 score (over 30 pages)

Essentially this number measures roughly how well a committee of four people could do on this task, by taking the most common answer selected by the four annotators.

The average committee-top-1 score from 1000 samples was 25.8

inter-annotator agreement is hard!

- too many phrases
- not enough overlap

measuring extraction quality

Top-1, Top-10

"To compute this measure, we counted the number of times the top output of our system for a given page was in the list of terms described by the annotator for that page. We divided this number by the maximum achievable top-1 score, and multiplied by 100."

system	top-1	top-10
MoC (Monolithic, Combined), *-Lin*	30.06^b	46.97^b
MoC (Monolithic, Combined), *All*	29.94	46.45
MoS (Monolithic, Separate), *All*	27.95	$44.13^‡$
DeS (Decomposed, Separate), *All*	$24.25^‡$	$39.11^‡$
KEA [7]	$23.57^‡$	$38.21^‡$
MoC (Monolithic, Combined), *IR*	$13.63^‡$	$25.67^‡$
MoC (Monolithic, Combined), *TFIDF*	$13.01^‡$	$19.03^‡$

Table 1: Performance of different systems

features		top-1	top-10	entropy
A	all	29.94^{b}	46.45^{b}	0.0113732^{b}
-C	capitalization	30.11	46.27	$0.0114219^{\dagger}$
-H	hypertext	30.79	$45.85^{\dagger}$	0.0114370
-IR	IR	$25.42^{\ddagger}$	$42.26^{\ddagger}$	$0.0119463^{\ddagger}$
-Len	length	30.49	$44.74^{\dagger}$	$0.0119803^{\ddagger}$
-Lin	linguistic	30.06	46.97	$0.0114853^{\ddagger}$
-Loc	location	29.52	$44.63^{\dagger}$	$0.0116400^{\ddagger}$
-M	meta	30.10	46.78	$0.0113633^{\ddagger}$
-Ms	meta section	29.33	46.33	0.0114031
-Q	query log	$24.82^{\dagger}$	$42.30^{\ddagger}$	$0.0121417^{\ddagger}$
-T	title	28.83	46.94	0.0114020
-U	URL	30.53	46.39	0.0114310

lower is better

Table 3: The system performance by removing one set of features in the MoC framework

Mapping Contextual Advertising to Sponsored Search

- Extract phrases from the publisher's web page
- Select ads using exact or advanced match on this phrase

conclusion

Approach based on logistic regression trained on editorial judgments

- Editors extracting salient terms from pages

Combining the information from multiple occurrences and treating the phrases as single units yields best results

IR and query log features account for almost all of the signal

Low precision – difficult problem

IR methods for content match

Berthier Ribeiro-Neto, Marco Cristo, Paulo B. Golgher, and Edleno Silva de Moura. 2005. Impedance coupling in content-targeted advertising. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). ACM, New York, NY, USA, 496-503.

The phrase extraction approach uses one feature of the page (phrase) to select the ads

- Risk with ambiguous phrases: 'Tahoe' is a destination as well as a truck model.

multi-feature selection

Can we select ads based on multiple features from the page?

What are the features of the ad?

How to weight the features?

What metrics to use to relate the ads to the pages?

$$a = \{w_{1,a}, w_{2,a}, \ldots, w_{n,a}\}$$

a is the visible part of the ad (title and abstract)

weights are TF-IDF

$$p = \{w_{1,p}, w_{2,p}, \ldots, w_{n,p}\}$$

the page p is a vector in the same space

formalism

$$\cos(\overrightarrow{p}, \overrightarrow{a}) = \frac{\overrightarrow{p} \cdot \overrightarrow{a}}{|\overrightarrow{a}||\overrightarrow{p}|} = \frac{\sum_k a_k p_k}{\sqrt{\sum_j a_j^2}\sqrt{\sum_j p_j^2}}$$

use cosine similarity

$$\texttt{AD}(p, a_i) = \mathrm{sim}(p, a_i)$$ based on visible parts of the AD

$$\texttt{KW}(p, a_i) = \mathrm{sim}(p, k_i)$$ based on keywords associated with the AD

$$\texttt{AD} - \texttt{KW}(p, a_i) = \mathrm{sim}(p, a_i \cup k_i)$$ based on keywords and the AD

basic set of measures

$$\texttt{AND} - \texttt{KW}(p, a_i) = \begin{cases} \mathrm{sim}(p, a_i \cup k_i) & \text{if} \quad k_i \subseteq p \\ 0 & \text{otherwise} \end{cases}$$ inclusion criteria

$$\texttt{AAK}(p, a_i) = \begin{cases} \mathrm{sim}(p, a_i \cup k_i) & \text{if} \quad a_i \cup k_i \subseteq p \\ 0 & \text{otherwise} \end{cases}$$

Language and the topic of the page and the ad can differ substantially:

- Publisher page belongs to a broader/ narrower contextual scope
- Ads concise in nature
- 'Hidden topic' – not mentioned in the ad and/or the page

The "impedance" problem

Intersection of the vocabularies of related pages and ads can be low: vocabulary impedance problem

impedance coupling

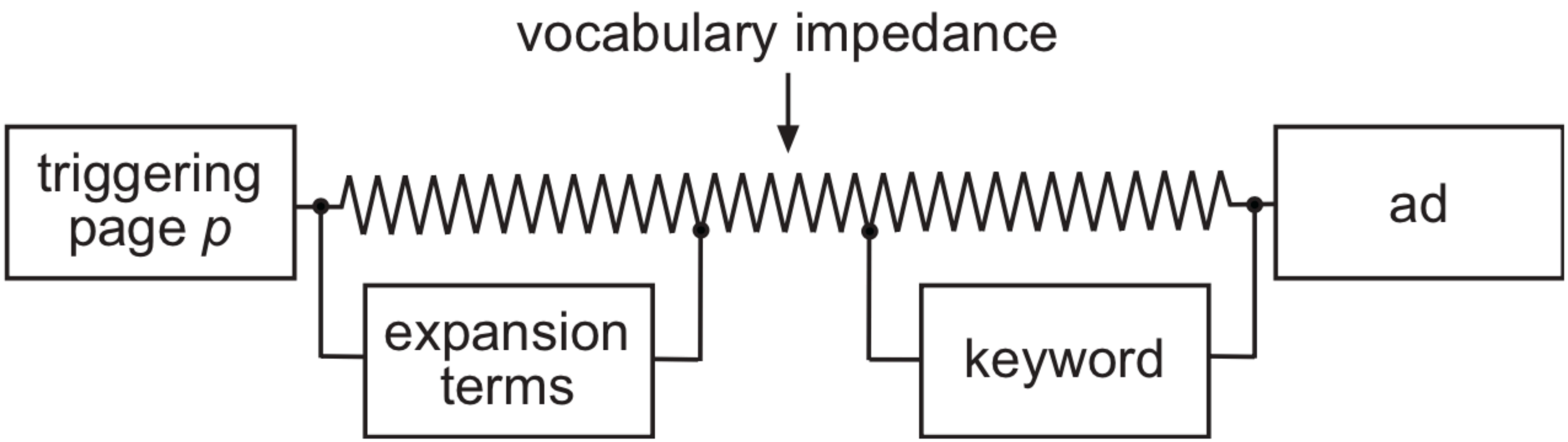


Figure 2: Addition of new terms to a Web page to reduce the vocabulary impedance.

a Bayesian formulation

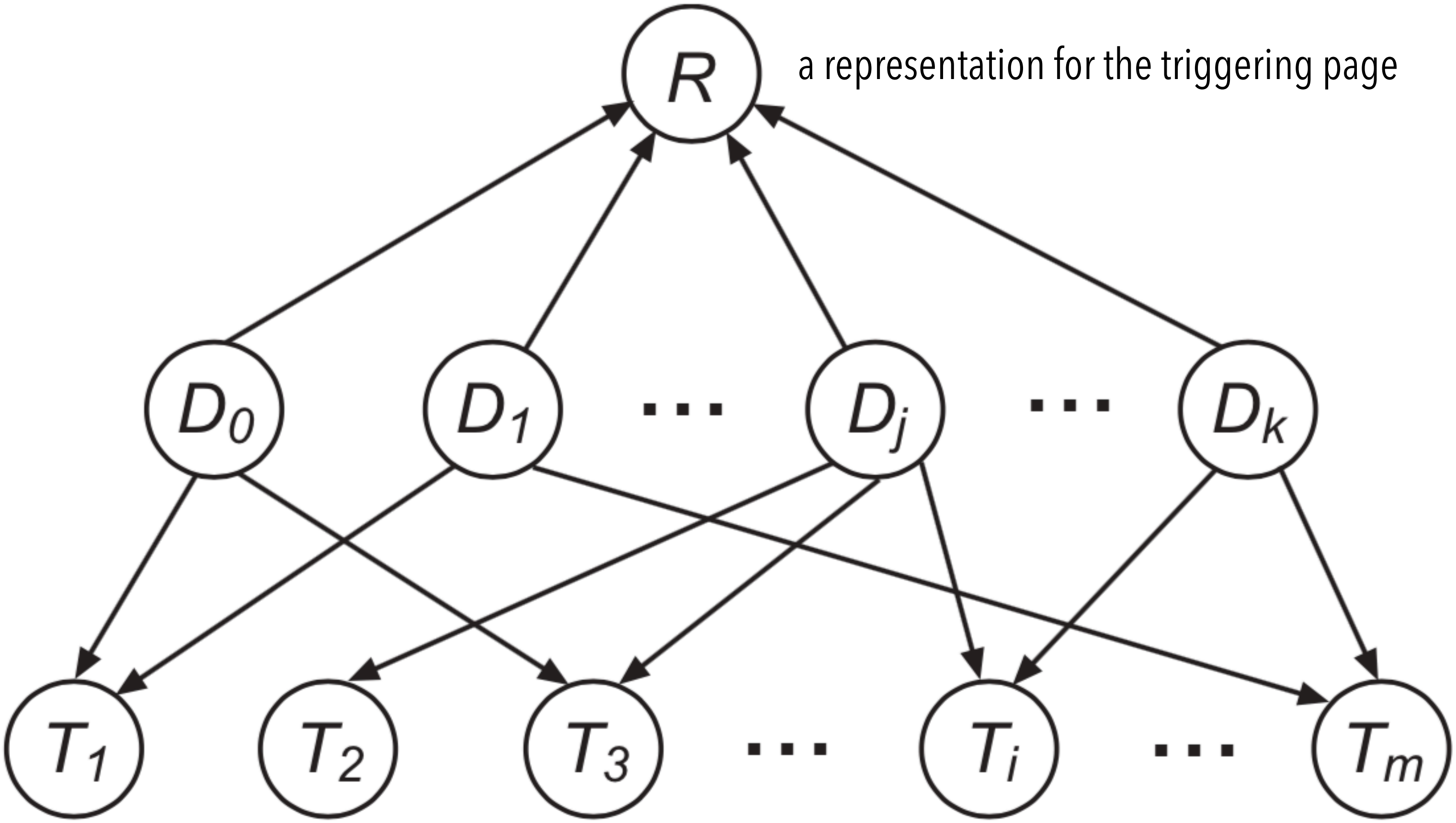
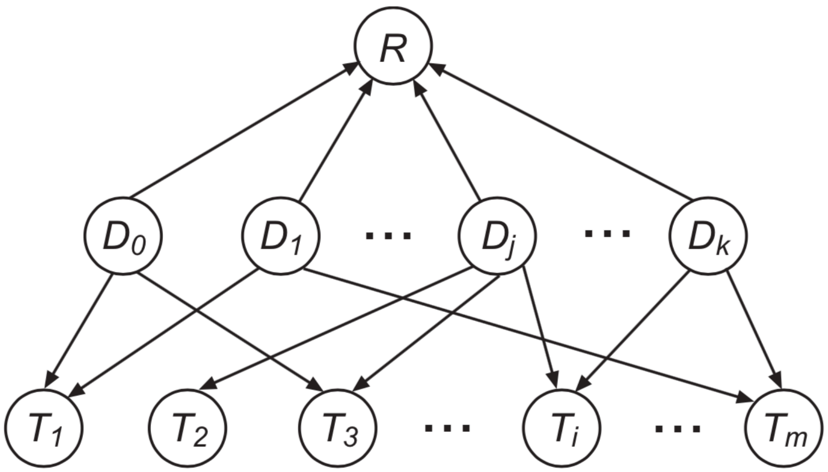


Figure 3: Bayesian network model for our impedance coupling technique.

$$P(T_i \mid R) = \frac{1}{P(R)} \sum_i P(T_i, R \mid d) P(d)$$

a Bayesian formulation



$$P(T_i \mid R) = \frac{1}{P(R)} \sum_i P(T_i, R \mid d)P(d)$$

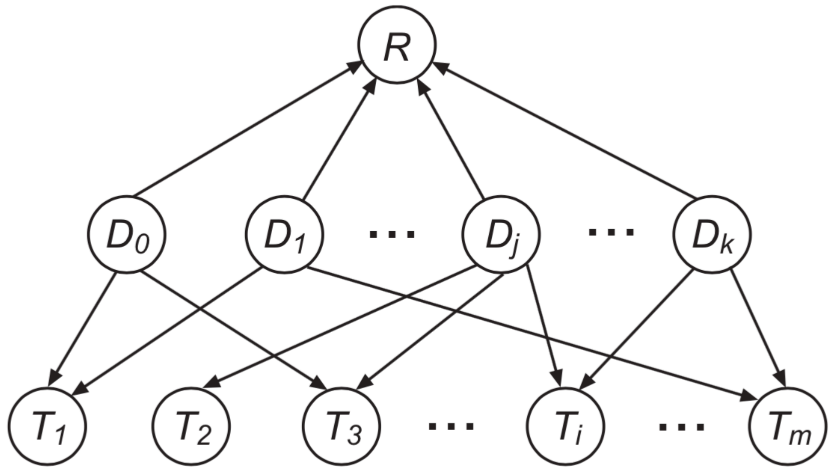
$$P(T_i \mid R) = \frac{1}{P(R)} \sum_i P(T_i \mid d)P(R \mid d)P(d)$$

$$P(T_i \mid R) = \frac{\nu}{P(R)} \sum_j P(T_i \mid d_j)P(R \mid d_j)$$

$$P(T_i \mid d_j) = \eta w_{i,j}$$

$$P(R_i \mid d_j) = \begin{cases} 1 - \alpha & j = 0 \\ \alpha \,\mathrm{sim}(r, d_j) & 1 \leq j \leq k \end{cases}$$

a Bayesian formulation



$$P(T_i \mid R) = \rho \left((1-\alpha) w_{i,0} + \alpha \sum_{j=1}^{k} w_{i,j} \, \mathrm{sim}(r, d_j) \right)$$

$$\frac{P(T_i \mid R)}{P(T_{\mathrm{top}} \mid R)} \geq \beta$$

$$\mathtt{AAK}(r, a_i), \mathtt{AAK}(p \cup r, a_i)$$

results: it works

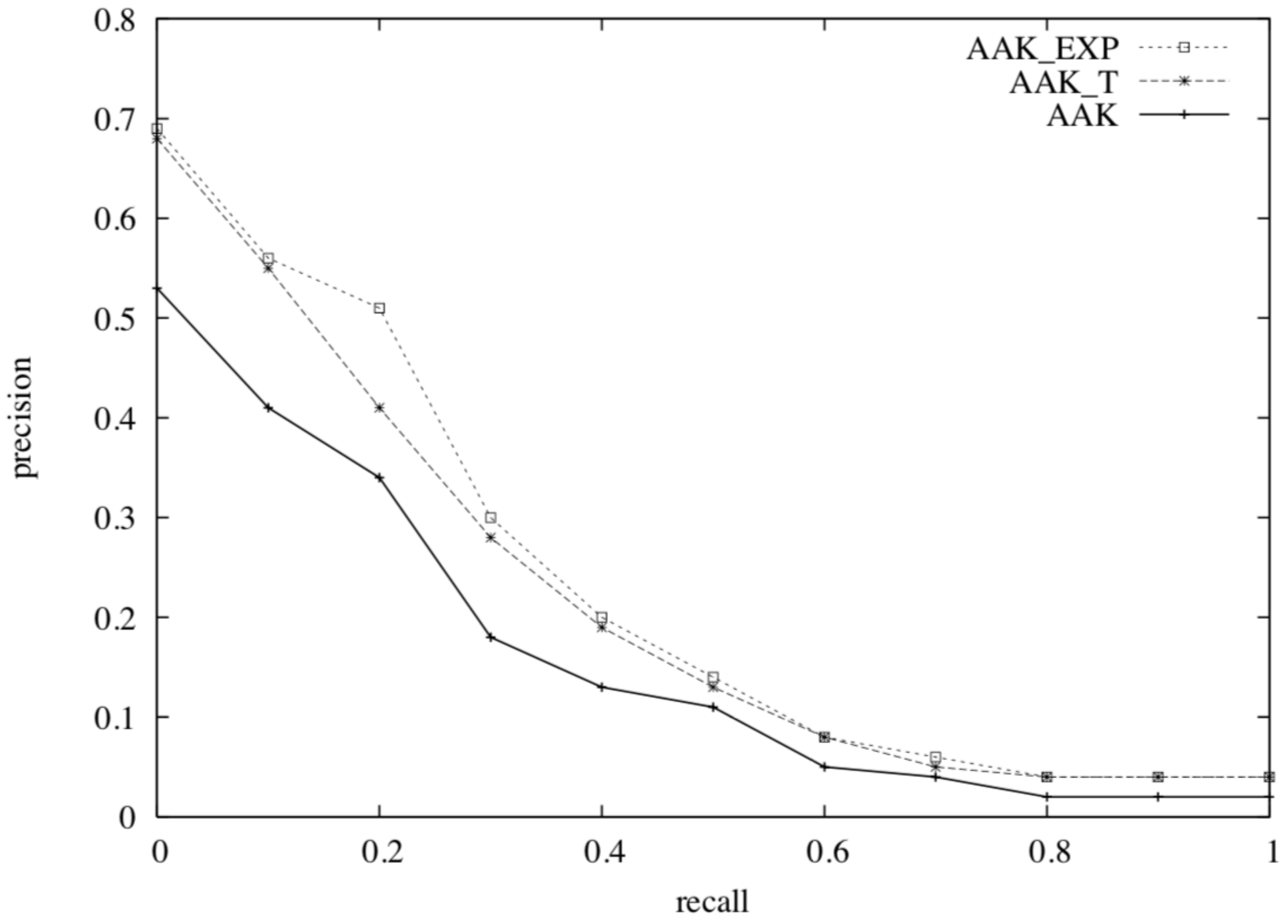


Figure 6: Impact of using a new representation for the triggering page, one that includes expansion terms.

results: it works

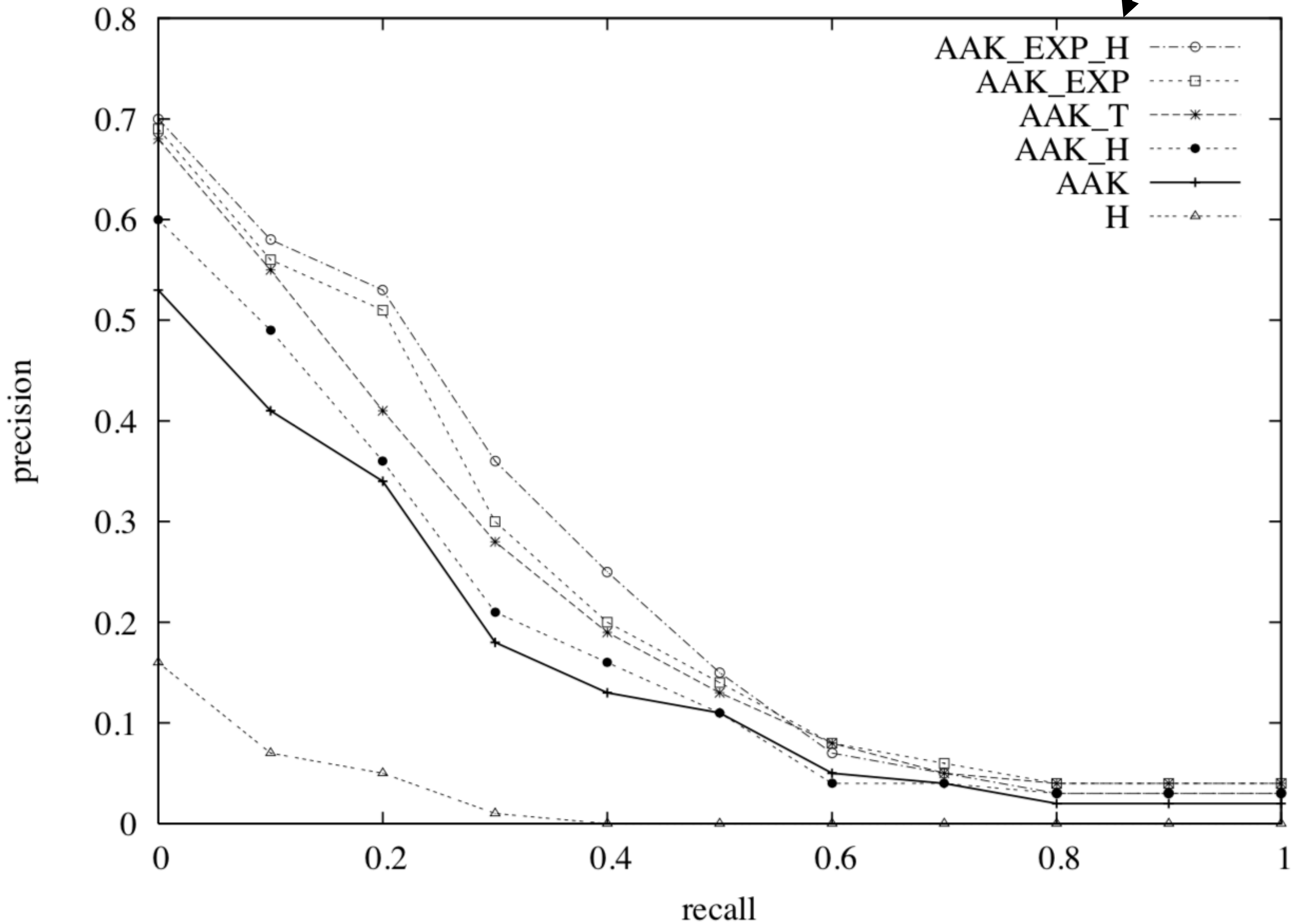


Figure 8: Comparison among our ad placement strategies.

Using IR techniques to match ads and pages

Both the ad and the page are mapped to a common vector space

Cosine of the angle between the ad and the page as the basic similarity measure

- Bid phrase as a required feature—projection of the space

Expanding pages using terms from similar pages improves results

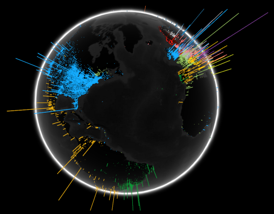
Landing page contains useful data for ad selection

summary

Some practical considerations:

- How long are the queries?
- How much is the cost of this method?

Introduction

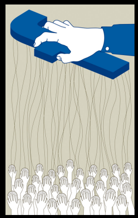


Web search

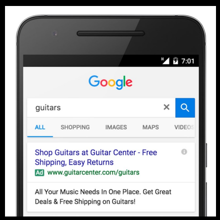
Game Theory

Auctions

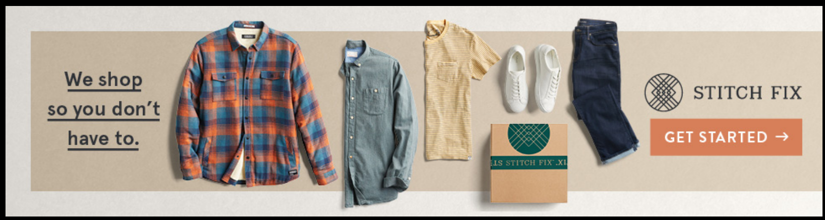
Data flows



Privacy



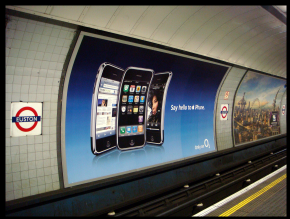
Text Ads



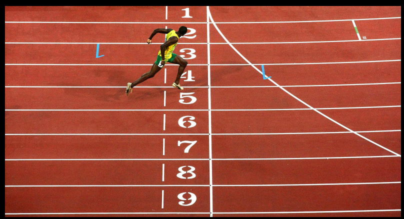
Contextual & Display Ads

Recommender systems

Behavioral targeting



Emerging areas



Final Presentations