

【SIGIR2013勉強会】

Session 7: Users and Interactive IR II

担当：梅本・佃（京都大学）

論文リスト

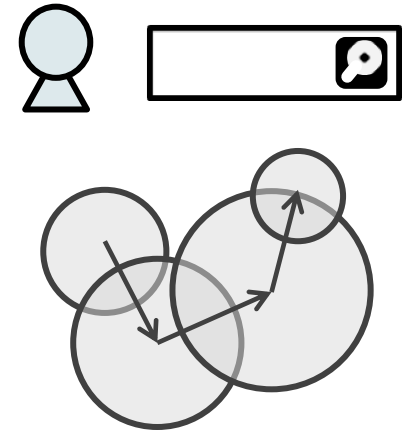
- 1. Looking Ahead: Query Preview in Exploratory Search**
 - ▶ Pernilla Qvarfordt, Gene Golovchinsky, Tony Dunnigan (FX Palo Alto Laboratory), Elena Agapie (Harvard University)
- 2. News Vertical Search: When and What to Display to Users**
 - ▶ Richard McCreadie, Craig Macdonald, Iadh Ounis (University of Glasgow)
- 3. Fighting Search Engine Amnesia: Reranking Repeated Results**
 - ▶ Milad Shokouhi (Microsoft), Rhen White, Paul Bennett (Microsoft Research), Filip Radlinski (Microsoft)
- 4. Toward Self-Correcting Search Engines: Using Underperforming Queries to Improve Search**
 - ▶ Ahmed Hassan, Ryen W. White, Yi-Min Wang (Microsoft Research Redmond)

Looking Ahead: Query Preview in Exploratory Search

Pernilla Qvarfordt, Gene Golovchinsky, Tony Dunnigan (FX Palo Alto Laboratory), Elena Agapie (Harvard University)

▶ Exploratory Search (探索型検索)

- ▶ 曖昧な情報要求・複雑な検索タスク
- ▶ 正解が1つでない・定まっていない
- ▶ 意思決定が絡む
 - ▶ 例) 京都観光の計画, 特許文書の検索, ...
- ▶ さまざまな観点で横断的・反復的に検索を行う必要



▶ Exploratory Searchを行う上での問題点

- ▶ recall-orientedなタスクでは, 多くの検索結果を閲覧
 - ▶ どの文書が既読／未読なのかをユーザは**覚えていない**
- ▶ 表示された検索結果を眺めることで既読／未読を判断
 - ▶ **クエリ入力時点**では, どの程度の既読文書があるのかがわからない

探索型検索用の提案インタフェース

基本的なアイデア

クエリ入力時点（検索結果表示前）で
既に表示／閲覧されたことのある文書を可視化

Query Preview



クエリ入力が300ms停止したり，文書が閲覧されたりした時にプレビューを更新

仮説と実験

Query Preview利用者の行動に関する3つの仮説

1. クエリ修正の際のユーザの**注意や行動**に影響を与える
2. 多くの**多様なページ**が得られるようなクエリを生成するようになる
3. 検索結果ページ中の**下位の文書**まで見るようになる

システム

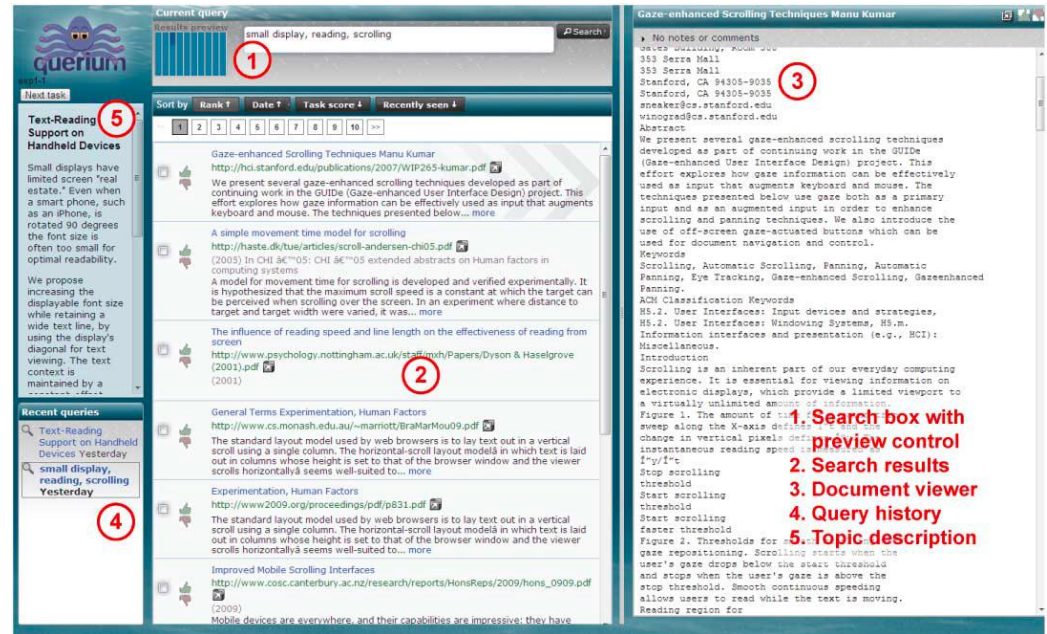
- ▶ Query Previewあり
 - ▶ preview condition
- ▶ Query Previewなし
 - ▶ control condition

タスク

- ▶ 特許申請をするために、できるだけ多くの関連論文を探す(6種類)

被験者

- ▶ 13人(タスク内容の知識あり)



実験結果

クエリ修正時のユーザの視線移動

- ▶ クエリ入力中は入力前後に比べ、画面を見ていない(クエリ作成に集中してるから?)
 - ▶ Previewシステムでは、入力中も画面が見られている
- ▶ 入力中に注目されるUI要素
 - ▶ クエリ部はPreview > Control, 結果部はPreview < Control, 他はどちらのシステムでも同等

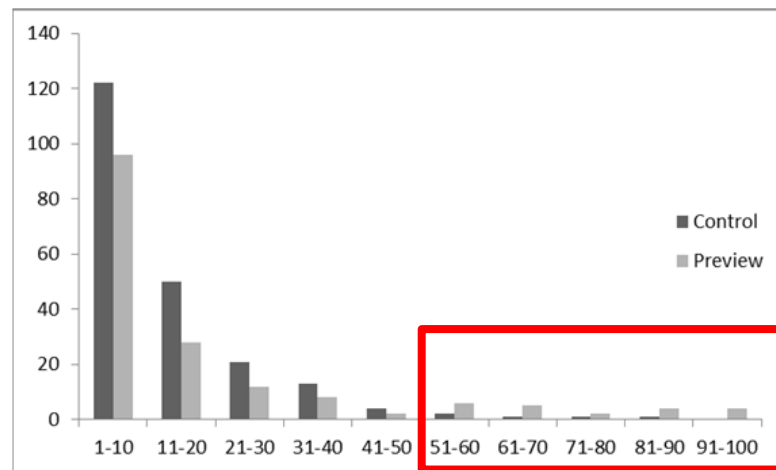
得られる結果の多様性

	Control	Preview
所要時間(MAX15分)	12.2	11.7
クエリ数	7.7	6.4
検索文書数	525	522
閲覧スニペット数	76.9	73.4
閲覧文書数	5.4	4.4
保存文書数	5.6	6.4

Previewシステムでは、**少ないクエリ数**で、(ほぼ)同数の文書を検索、**多くの文書**を保存

保存された文書の再現率(および適合率)もPreviewシステムが上回る

下位結果に対する閲覧



Previewシステムでは、**下位文書の閲覧割合が高い**(スニペット閲覧率も同傾向)

News Vertical Search: When and What to Display Users

Richard McCreadie, Craig Macdonald, Iadh Ounis (University of Glasgow)

▶ Vertical Searchとは？



- 特定ソースからの検索結果
 - ▶ ニュース, 画像, 動画, ...
 - ブロックとして検索結果に挿入
- ニュース検索結果の情報源は電子新聞が大部分

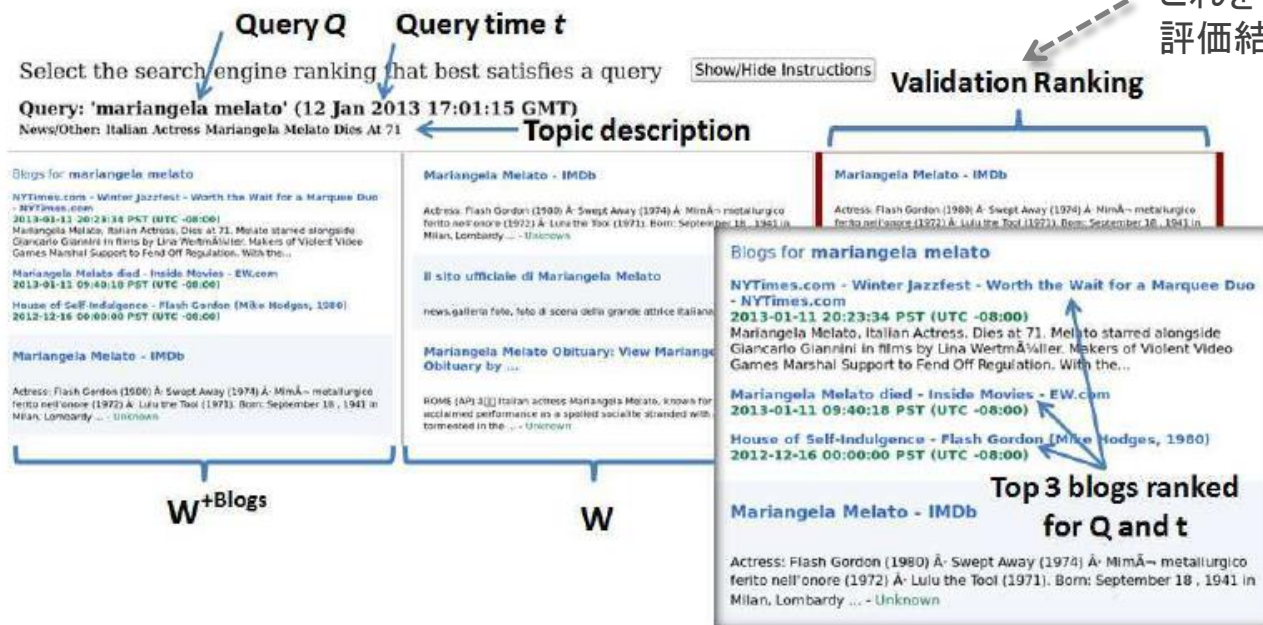
▶ 本研究で取り上げる3つのResearch Questions

- ▶ [RQ1] ニュース情報の**集約源**は**電子新聞**だけで十分？
- ▶ [RQ2] **User-Generated-Content**はニュース検索に有用？
- ▶ [RQ3] イベント発生からの**経過時間**と有用な情報源との関係？

News-verticalの分析手法

- ▶ 通常のWeb検索結果 W とニュース情報集約結果 W^{+e} を比較 ($e \in \{News, Blogs, Twitter, Wikipedia\}$)
- ▶ クラウドソーシングサービスでユーザ実験

ダミーデータ
これをクリックした場合、
評価結果からデータを除外



- ▶ アンケートの内容
 - ▶ W^{+e} の選択・非選択理由を回答
 - ▶ 適合性, 新鮮度, 冗長性, ...

ユーザ実験用のデータセットの作成

▶ 異なる時期の2種類のデータセット

- ▶ 2012_{Apr}: 2012年4月11日～23日のニュースクエリ199個
- ▶ 2013_{Jan}: 2013年1月10日～16日のニュースクエリ98個

▶ ニュース関係のクエリの抽出源

- ▶ 2012_{Apr}: Google Trendsで1時間ごとの人気キーワード
- ▶ 2013_{Jan}: Bitly Data APIで30分ごとのBursting Phrases

▶ 検索結果の作成

- ▶ Bing APIでWeb検索結果
- ▶ 関連するNews3件, Blogs3件, Tweets3件, Wikipedia1件を集約

▶ 各検索結果対を最低10人に評価

- ▶ 合計で98人が参加(かかった費用は217.25ドル)

実験結果

[RQ1] W vs. W^{+News}

- ▶ News集約結果の選択が48.5%
 - ▶ 約半数のニュース関連クエリは、新聞記事は有用でない
- ▶ 選択理由
 - ▶ クエリに適合: 77.2%
 - ▶ イベントに関連する記事: 63.3%
 - ▶ 速報内容が記述されている: 28.2%

[RQ2] UGCのニュース関連クエリに対する有用性

- ▶ 各UGC集約結果の選択率
 - ▶ Blogs: 56.9%
 - ▶ **Twitter**: 71.3%
 - ▶ Wikipedia: 52.1%
 - ▶ UGC間の選択率の比較
 - ▶ Blogs > Twitter: 22.5%
 - ▶ Blogs < Twitter: 48.5%
- **各UGCには得意・苦手分野がある**

[RQ3] イベント経過時間と有用な情報源との関係

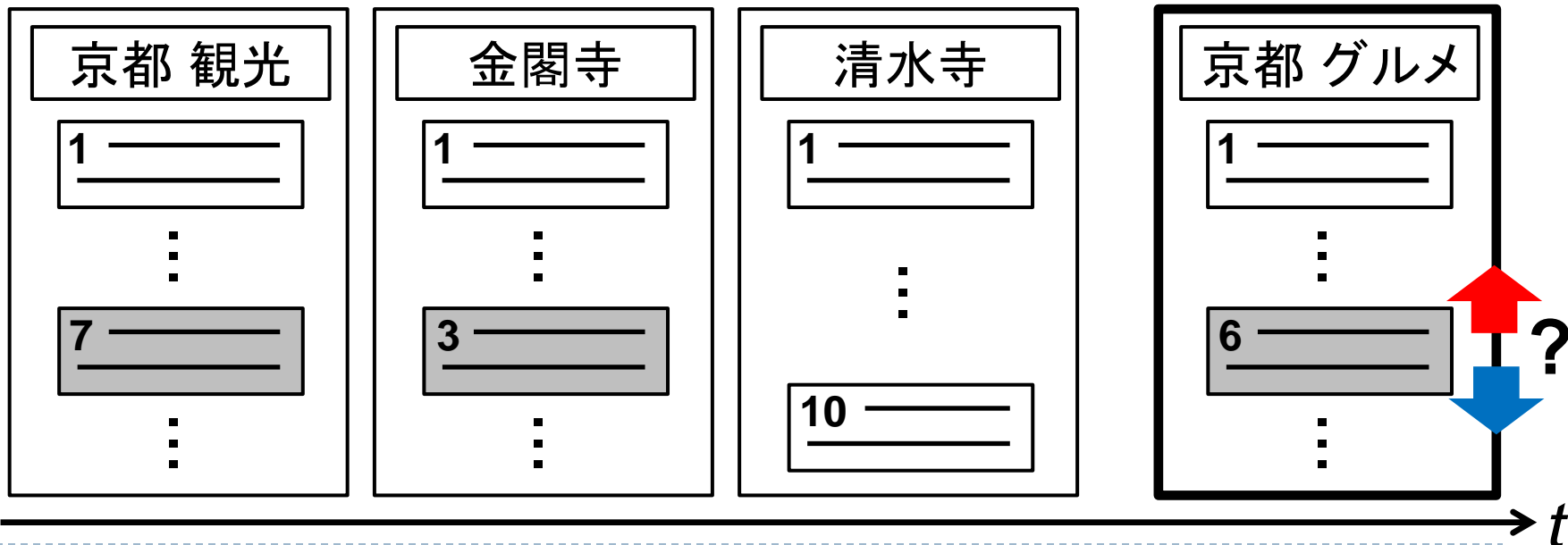
- ▶ ニュースクエリを3種類に分割
 - ▶ Breaking: 発生から12時間以内
 - ▶ Recent: 12時間～48時間
 - ▶ Long-Running: それ以上
- ▶ タイプごとに有用性を分析
 - ▶ Newsの有用性は時間とともに減衰
 - ▶ Blogsは発生時間からやや遅れて有用に
 - ▶ Twitterは**最新情報**の取得に有用

Finding Search Engine Amnesia: Reranking Repeated Results

Milad Shokouhi (Microsoft), Rhen White, Paul Bennett (Microsoft Research), Filip Radlinski (Microsoft)

概要

- ▶ 1つの検索セッション中で入力される複数のクエリに対して同じページが検索結果に提示されることがある
- ▶ 複数回提示されるページでも、そのセッション中でのそのページに対するアクション(クリックしたか等)に応じて、そのページをより上位または下位にリランキングすることが目的



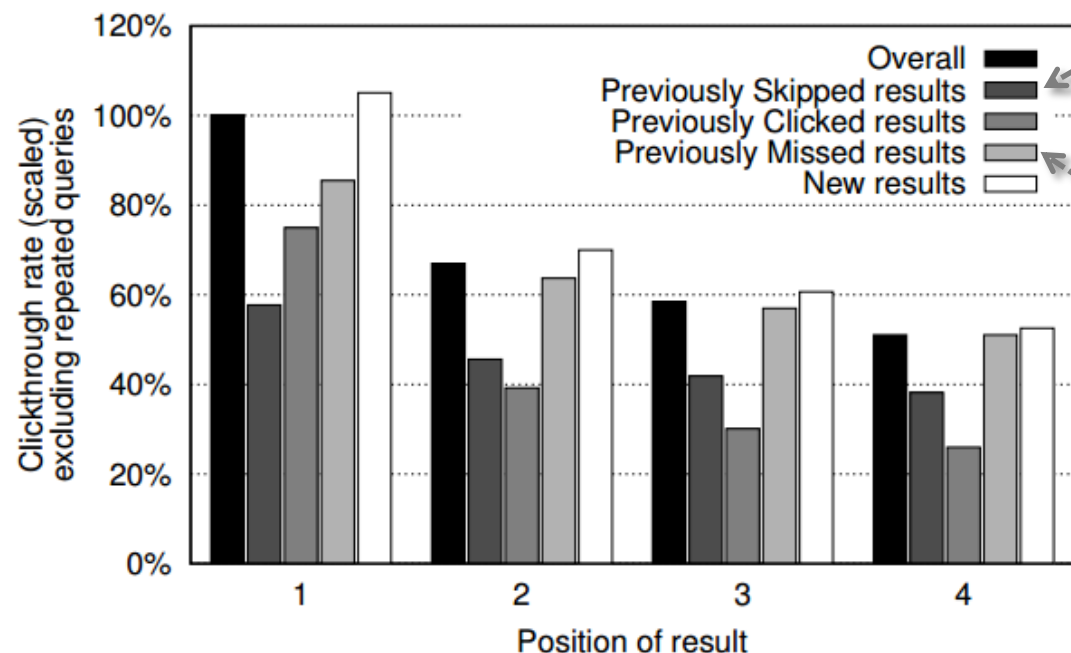
データ分析 (BingとYandexのクエリログ使用)

- ▶ 1検索セッション中に同一ページが複数回出現する割合
 - ▶ 2クエリ/セッションしかない場合でも44%のセッションが該当
 - ▶ 9クエリ/セッションの場合は90%以上のセッションが該当
- ▶ 同一セッション内で複数提示されるページに対する反応は時と場合によって異なるか。

過去にそのページが提示されたときにそれより下位がクリックされた

過去にそのページが提示されたときにそれより下位がクリックされなかった

過去に見落とされたページは複数回現れてもクリックされる可能性大



ページの重要度推定

- ▶ LambdaMARTを使用
- ▶ 各ページの素性は以下の4カテゴリから成る
 - ▶ Click features
 - ▶ 同一セッション内でこれまでにクリックされた回数など
 - ▶ Display features
 - ▶ 同一セッション内でこれまでに検索結果に出現した回数など
 - ▶ Query features (ページ非依存)
 - ▶ 今入力しているクエリと, 同一セッション内でこれまでに入力したクエリとの平均類似度など
 - ▶ Other features
 - ▶ 検索エンジンによって計算されるページのスコアなど

実験

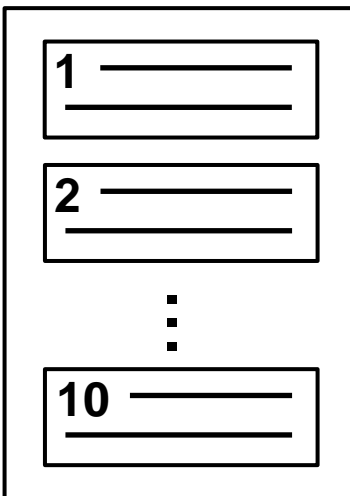
- ▶ セッションの一番最後にクリックされたページと、滞在時間が30秒以上のページをユーザが満足したページ(正解ページ)として、リランキング前後のMRRとMAPを計算
- ▶ MRR
 - ▶ リランキング前: 0.696 リランキング後: 0.711 (2.1%改善)
- ▶ MAP
 - ▶ リランキング前: 0.536 リランキング後: 0.553 (3.2%改善)
- ▶ 順位が変動したページのうち、63%は順位が1つだけ上昇または下降したものであった

全体像

red cliff 🔍

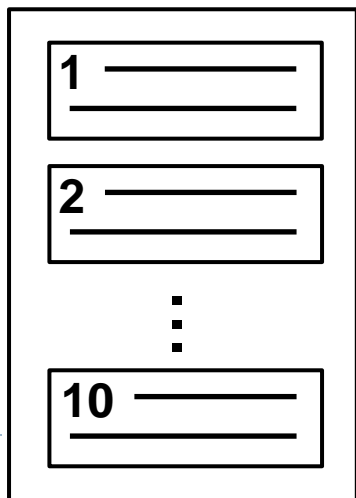
$(1, 0, 0, 1, 0, \dots, 1)$

オリジナル



ranker

再ランキング



クエリの各クラスに対して
推定されたSAT・DSAT

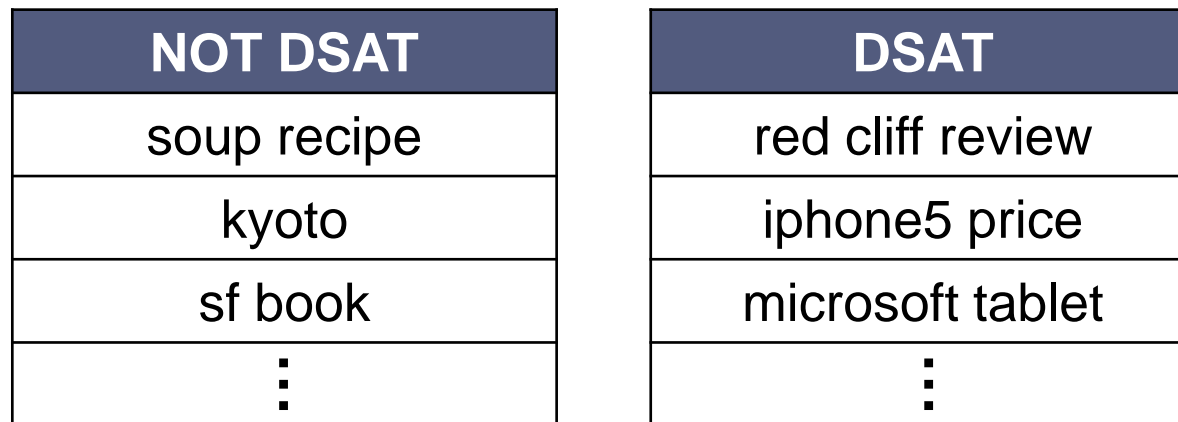
SAT
$(0, 0, 1, 0, 0, \dots, 1)$
$(1, 0, 0, 0, 0, \dots, 0)$
$(0, 0, 0, 1, 0, \dots, 1)$
$(1, 1, 0, 0, 1, \dots, 0)$

DSAT
$(0, 1, 1, 0, 0, \dots, 1)$
$(1, 0, 0, 1, 0, \dots, 1)$
$(0, 0, 0, 1, 1, \dots, 0)$
$(1, 1, 0, 1, 0, \dots, 0)$

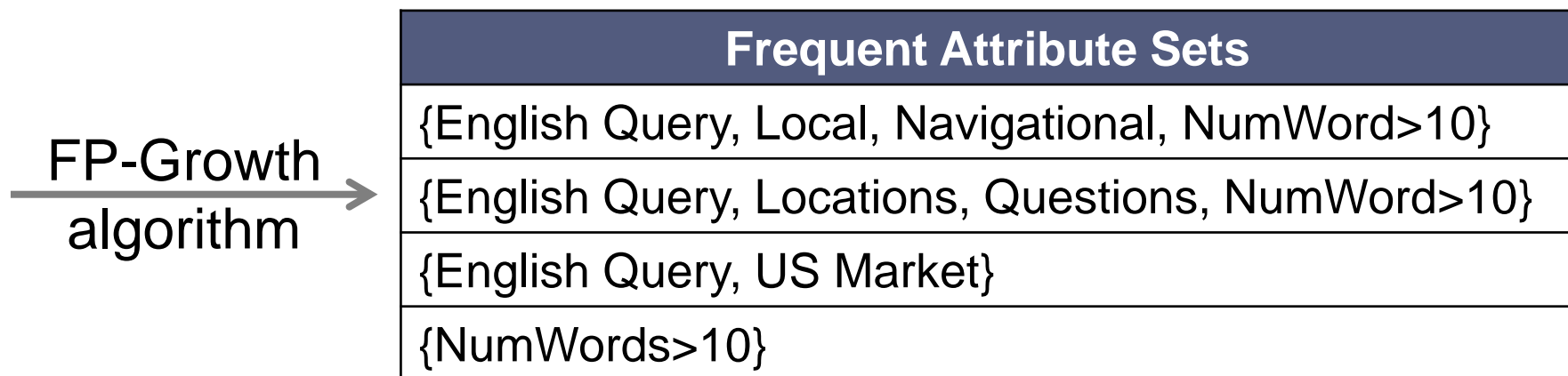
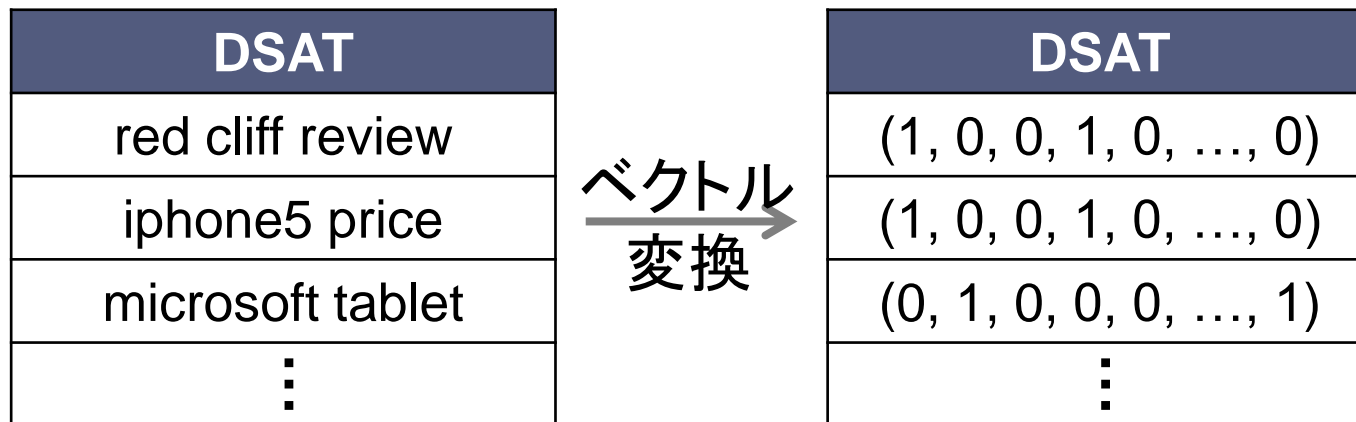
クエリのクラスのSAT・DSAT推定

検索セッション中で検索エンジンの切り替えがあったクエリ
“soup recipe”, “red cliff review”, “kyoto”, “sf book”,
“iphone5 price”, “microsoft tablet”, “lambda mart”, ...

既存の分類器



DSATに頻出するクエリクラスの特徴



DSAT → (NumWords>10) は成立しても
(NumWords>10) → DSAT は不成立



両方向の確率が高い
パターンを発見する
必要あり

真にDSATなクエリのクラスの特定



- ・クリック数が閾値以上
- ・最後に30秒以上ページを閲覧



SAT
kyoto kinkaku-ji
steve jobs
sigir 2013
⋮

SAT
(1, 0, 1 1, ..., 0)
(1, 0, 0, 0, ..., 1)
(1, 0, 0, 0, ..., 0)
⋮

ベクトル →

FP-Growth algorithm →

Frequent Attribute Sets
{English Query, Local, NumWord>10}
{English Query, Locations}
{US Market}
{NumWords>10}

A={English Query, Locations} として

$$DSAT \text{ correlation} = \frac{P(A, DSAT)}{P(A)P(DSAT)}$$

が1.3以上のクエリを真のDSATとする