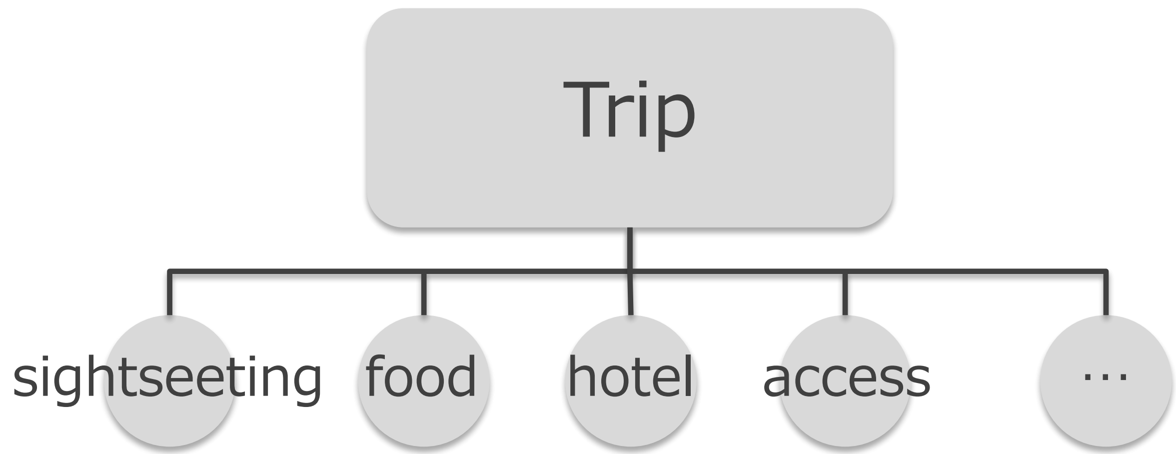
Predicting Next Query Reformulation Type from Current Search Behavior

Kazutoshi Umemoto (umemoto@dl.kuis.kyoto-u.ac.jp)

Tanaka & Tajima Laboratory
Dept. of Social Informatics, Graduate School of Informatics
(Kyoto University)

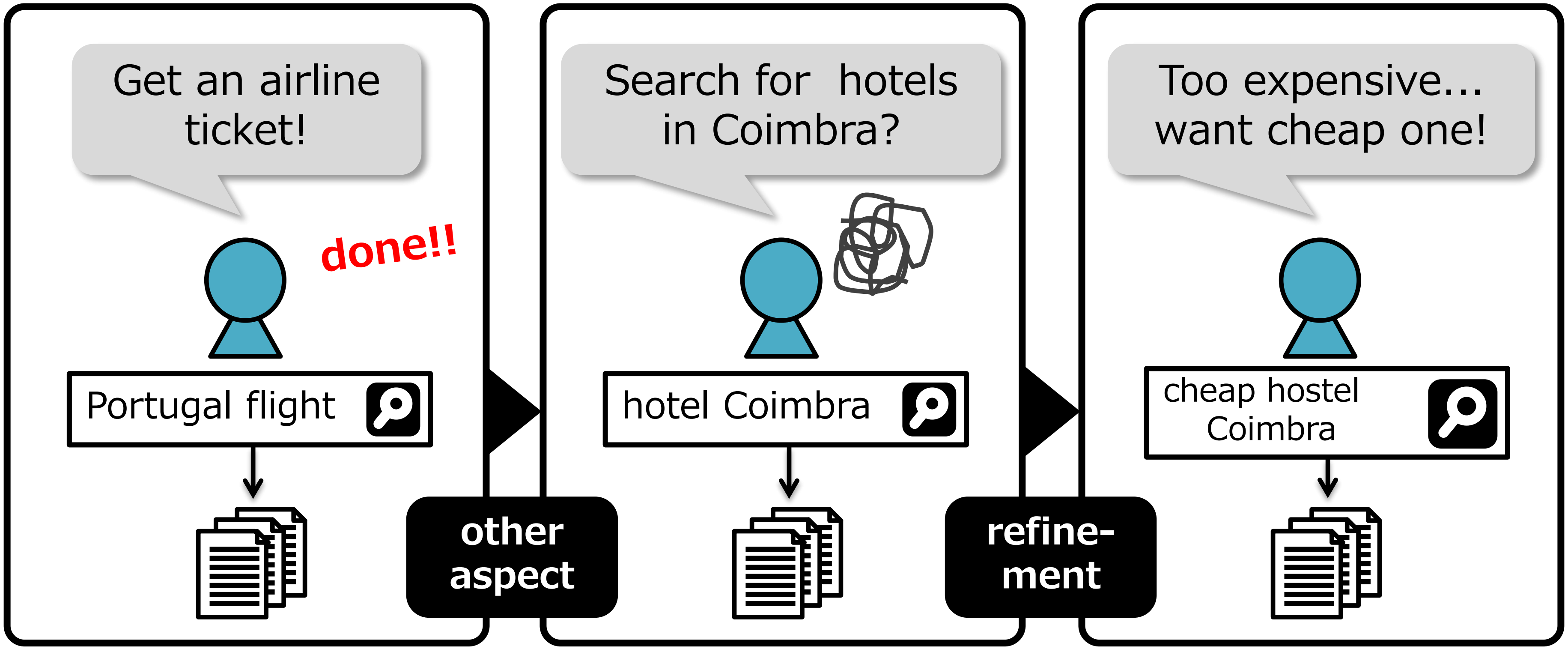
Tasks in Web Search

- **About 20% search tasks consist of multiple sub-goals** **[Jones CIKM'08]**
  - hard to meet information need with single query



Example of Complex Search Task

Search task about trip to Coimbra

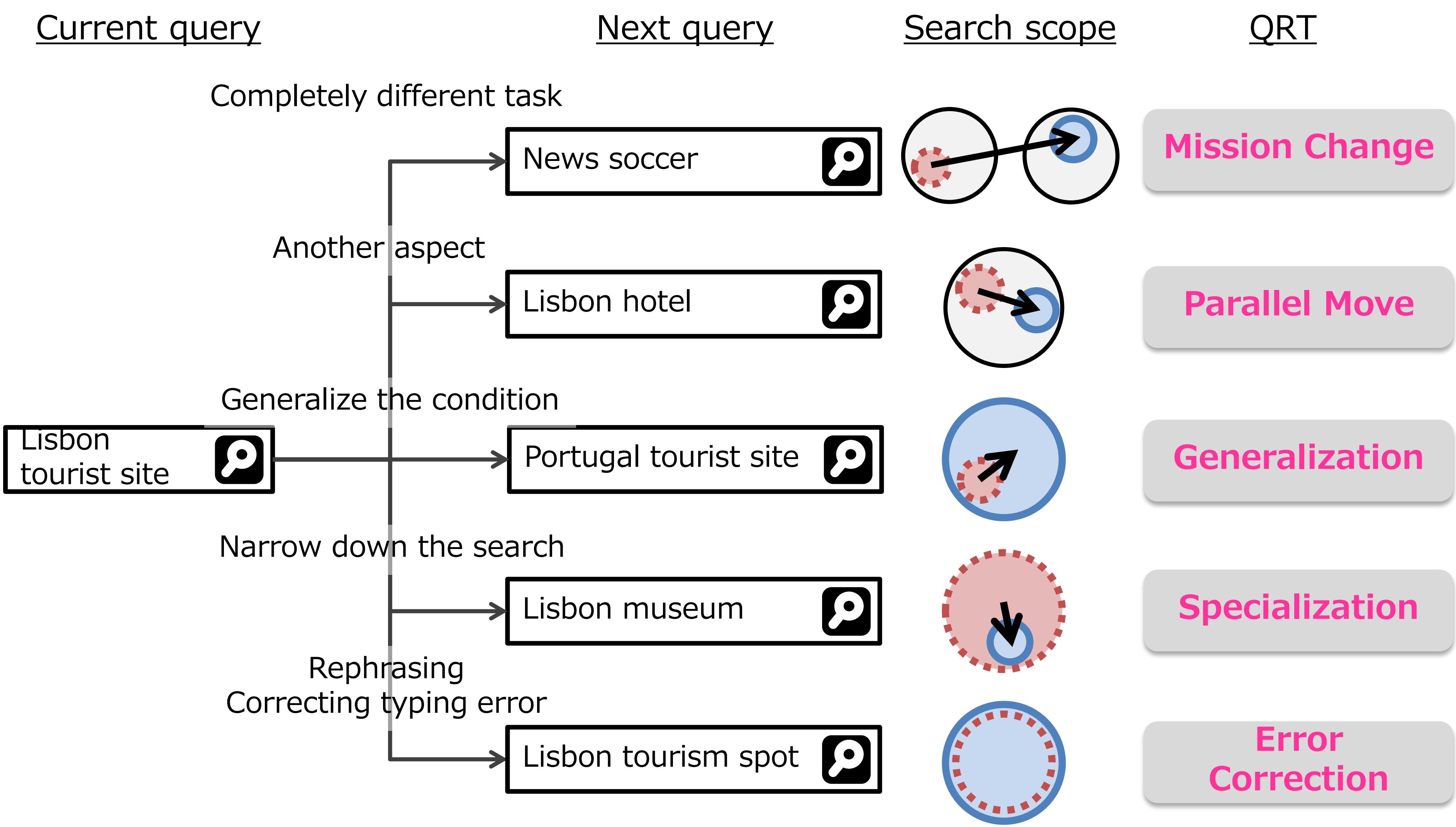


Characteristics

Reformulate search queries iteratively

in order to meet their sub-goals

Five Query Reformulation Types (QRT) [1]



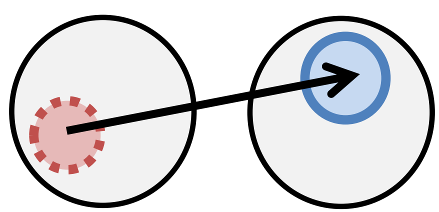
[1] P. Boldi et al. “From ‘dango’ to ‘japanese cakes’: Query reformulation models and patterns”, (WI '09)

1st Query Reformulation Type

Mission Change

Completely different task

Lisbon tourist site → News soccer



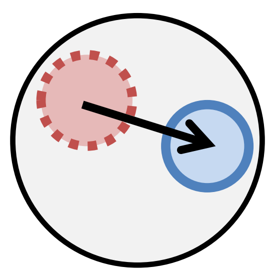
Change of Search Scope

2nd Query Reformulation Type

Parallel Move

Another aspect in the current task

Lisbon tourist site → Lisbon hotel



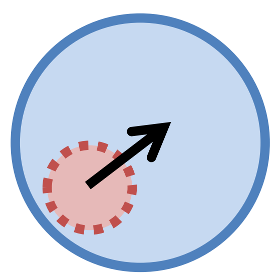
Change of Search Scope

3rd Query Reformulation Type

Generalization

Generalize the search condition

Lisbon tourist site → Portugal tourist site



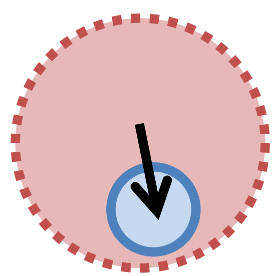
Change of Search Scope

4th Query Reformulation Type

Specialization

Narrow down the current search

Lisbon tourist site → Lisbon museum



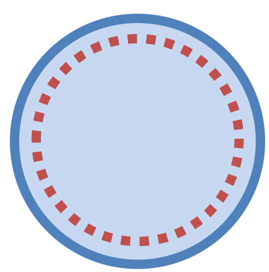
Change of Search Scope

5th Query Reformulation Type

Error Correction

Rephrasing OR spelling error correction

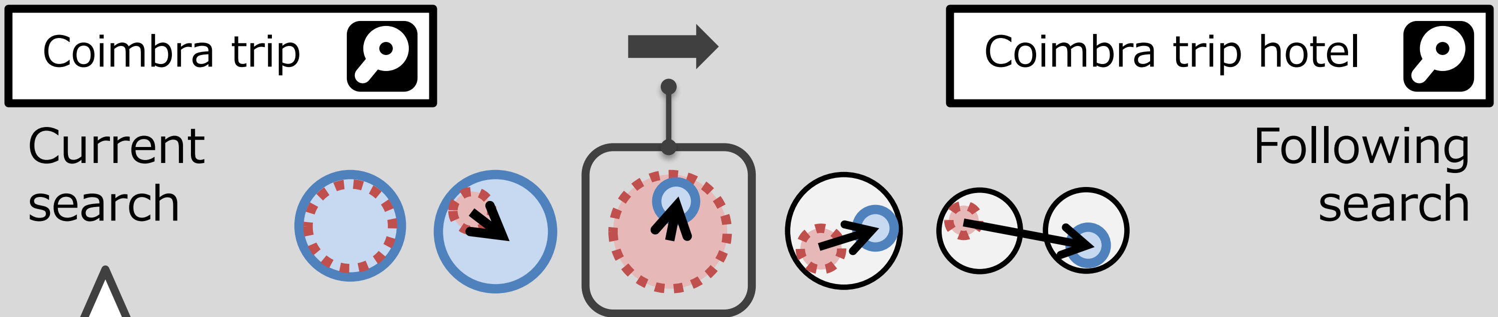
Lisbon tourist site → Lisbon tourism spot



Change of Search Scope

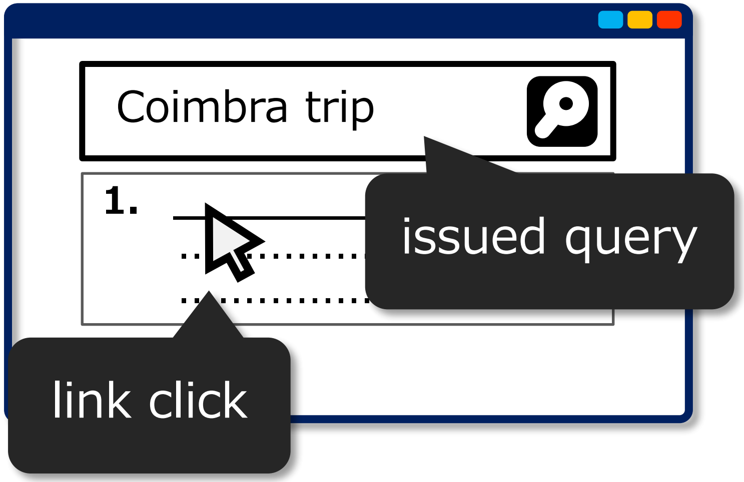
Research Purpose and Approach

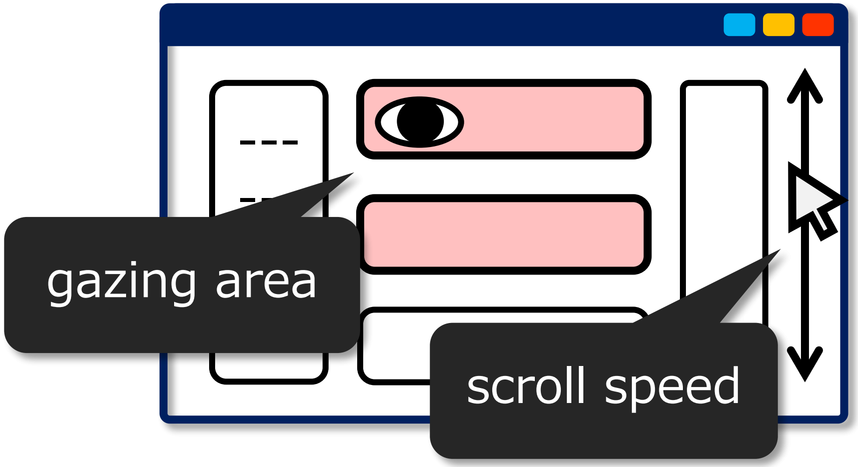
Predict the following QRT when finishing current search



User behavior could affect the following search

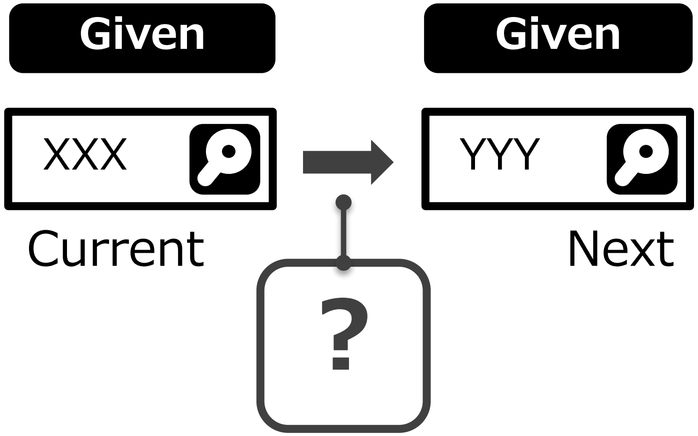
- search engine result page
- search trails





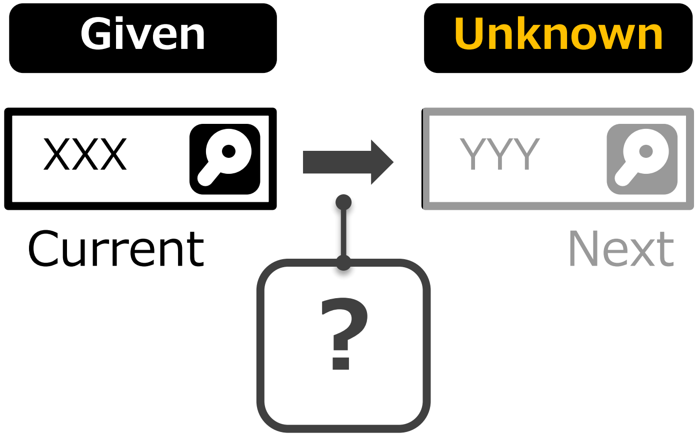
Related Work

- **Existing work [1]**
  - ▸ Estimate QRT between two queries
  - ▸ Required both search info.



[1] P. Boldi et al. "From 'dango' to 'japanese cakes': Query reformulation models and patterns", (WI '09)

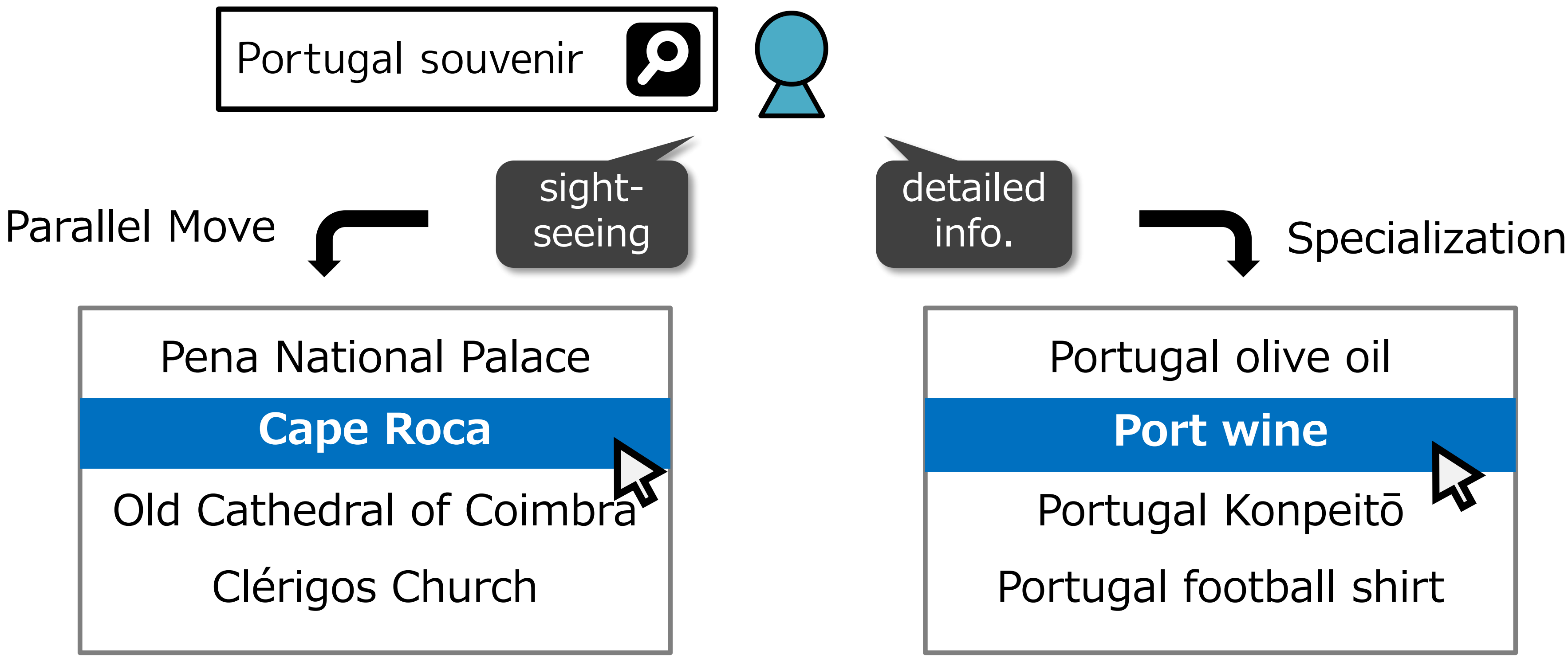
- **Our work**
  - ▸ Predict the next QRT
  - ▸ Use only current search info.



➡ applicable to **online search support**

Possible Application Example

Dynamic Query Suggestion based on User Behavior



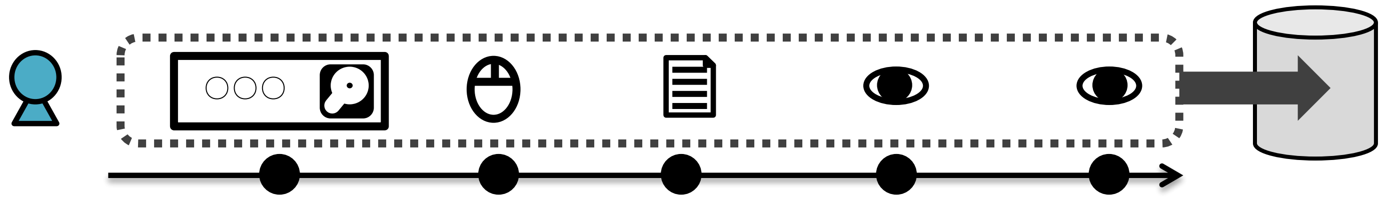
Change suggested queries according to predicted QRT

PROPOSED METHOD

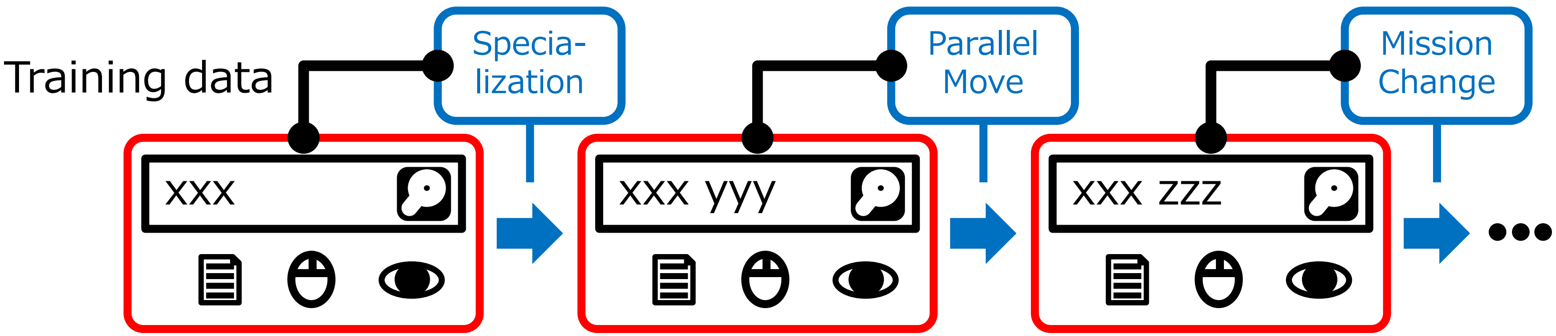
Research Flow

QRT classification by machine learning using behavior log data

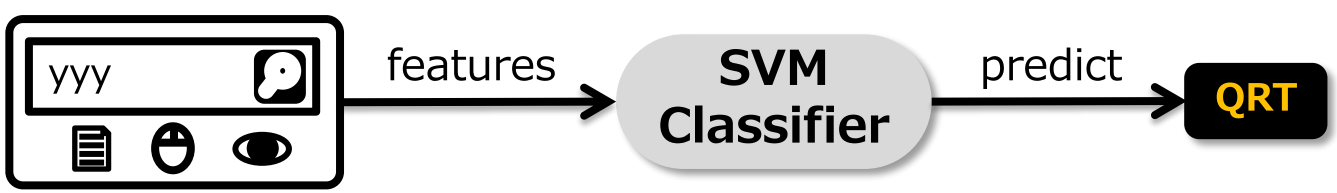
1 **Collecting behavior log data in Web search**



2 **Making training data from collected data**

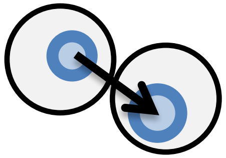


3 **Constructing QRT classifier from training data**



Possible Cause of Query Reformulation

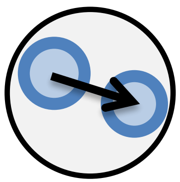
Mission Change



- Interest transition to another contents

➡ features about **user interest to current search**

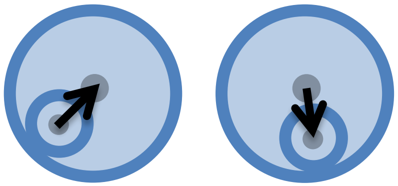
Parallel Move



- Obtained sufficient info. from current aspect

➡ features about **user commitment to current search**

Generalization / Specialization



- Too narrow/broad information

➡ features about **current search scope**

Features for QRT Classification

67 features from 6 categories

Transition Features

Previous search

- previous QRT

Query Features

Search Scope

- #{terms in query}
- hitcount

Span Features

User Commitment

- browse time
- #{browse pages}

Click Features

**(Explicit)
User Interest**

- #{clicked results}
- #{clicked ads}

Mouse Features

**(Explicit)
User Interest**

- scroll speed

Gaze Features

**(Implicit)
User Interest**

- gaze area
- attention for ads

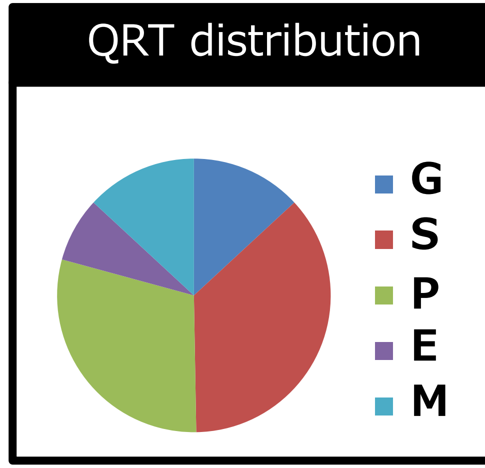
EVALUATIONS

Experimental Setup

<u>Objective</u>

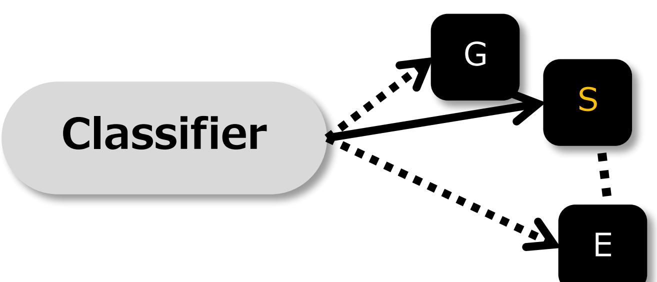
- To what extent can our method predict QRTs accurately?
- Which kinds of behavior contribute to QRT prediction?

- **Logged *my* search behavior**
  - Five days (4/4, 4/25, 5/4, 5/6, 5/7)
  - total # of queries = **183**



- **Construct two types of classifiers**

1. whole QRTs classifier



2. each QRT classifier

Whole QRTs Classification

- **Accuracy**
  - calculated through *K*-fold cross validation (*K*=183)

Predict QRT from user behavior	output QRT at random	always output most dominant QRT (Specialization)
Proposed	Baseline$_{rand}$	Baseline$_{dom}$
41.0%	20.0%	36.6%

- Proposed method ＞ Baselines

⇨ Additional training data may improve accuracy

Features Contributing to Whole QRTs Classification

Feature removal test

Accuracy decreases after removal ⇒ High contributing feature

○ High contributing feature

- click for search results
- gaze to left area in Web page
- #{terms in query}

✖ Low contributing feature

- last QRT is Generalization
- gaze to center area in Web page

Relevance of search results

Interest for non-main info.

broadness of Search scope

Features Contributing to Each QRT Classification

Generalization

Hitcount

↑

too few search results

Parallel Move, Mission Change

Ad-related features

↑

interest shift to another topic

Specialization

#{terms in query}

↑

broad results due to short query

Error Correction

#{characters of query}

↑

typing error for long query

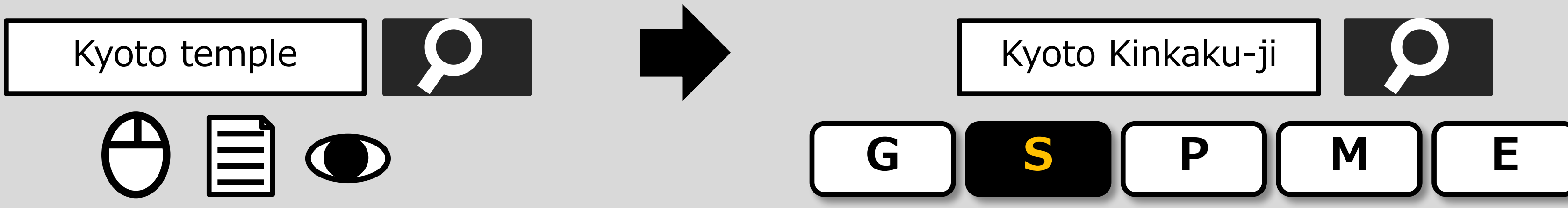
DISCUSSIONS

Considering the Past QRT Sequence

Approach in this work

1st order Markov model

Predict next QRT from current search info.



Need to consider search task info. for more accurate classification

HMM or CRF may be applicable to this problem

User Dependency for QRT Classification

this work

Predict next QRT using *only my* behavior data

Is my classifier applicable to other's behavior?

User independent??

- search scope for query
- relevance for results

User dependent??

- Mouse operation
- Eye movement

ToDo

Find out user dependency of features effective for QRT classification

Summary

Query Reformulation Type Prediction

Approach

- construct a classifier from user behavior log data (including query, mouse and eye movements)

Evaluation

- Classification accuracy: about 40%
- Contributing features: *Gaze position*, #{*Terms in query*}

Future work

- Consider applicability of HMM or CRF
- Find out user dependency of each feature