

# 学習型情報探索行動を探る ～CRES共同研究プロジェクトの成果から～

高久 雅生（物質・材料研究機構）



# 発表の概要

- ウェブ情報検索を中心とした情報探索を対象としてとりわけ学習型情報探索(Exploratory Search) をキーワードにして、これまで4年間にわたって実施してきたCRES共同研究プロジェクトにおけるウェブ情報探索行動の研究成果を報告する。くわえて、関連する分析手法やツールなど、ユーザ実験の方法論についても紹介する。

# お品書き

- はじめに
  - プロジェクトCRESの紹介
  - 研究の動機と目的
- 研究手法—ユーザ実験とその分析
- ユーザ実験における収集データ
- ユーザ実験事例
- 開発手法とその結果
  - ウェブ行動力テゴリ, 閲覧ページ分類, LinkDepth, コンセプトマップ, Lookzone, etc.
- 課題

# CRESプロジェクト <http://cres.jpn.org>

- CRES (Cognitive Research for Exploratory Search)
  - 2007年から6名で始めた共同研究プロジェクト
  - 月例研究会・MLを通じた議論
  - 実験・分析を通じた手法開発
- 認知科学の知見と情報検索研究の知見をむすぶ
- Web情報探索の利用行動そのものに着目
- 幅広く情報探索行動研究につかえる手法を開発

# 研究の動機

- NTCIR-5 WEBにおけるテストコレクション構築の経験
  - 判定者による揺れ、「情報ニーズ」「適合性」概念の難しさ
  - → もう少し「ひと」に近い部分の情報検索研究がしたい
- 1990-2000年代を通じた情報検索研究
  - 大量の文書データ・利用者行動データに基づいた研究がさかん
  - 情報検索のパラダイムのゆるやかな変化
  - 古典的情報検索システムのモデル **+ α**
    - 1)文書群, 2)利用者, 3)クエリー, 4)ランキングリスト, ...
    - 主題的適合性, ...
- 情報検索研究のトレンドから...
  - Human Interaction; メディアの多様性; ソーシャルメディア

# 研究の目的

- Web情報探索行動の理解

ひとはどのようにWebを利用しているか？

- 探索行動にかかる要因

タスクやユーザといった属性の違いは行動に  
どのような影響を与えるか？

- 包括的な行動データによる探索過程の分析

情報探索中にどのような行動をとるか？  
ページのどこに着目しているか？

# Web情報探索行動の理解

- 調査学習型探索 (Exploratory Search) [Marchionini, 2006]
  - 事項・事実検索
  - 調査や学習における探索
    - 探索のゴールを意識しながら
    - 新しい知識を獲得しながら
- タスク遂行中の情報利用行動
  - 事実発見、情報収集、ブラウジング、巡回、トランザクション [Kellar et al., 2007]
- 収集・分析データ
  - ブラウザログ（ユーザ行動）、視線データ、発話、アンケート、インタビュー

# ユーザ実験

- 実験参加者
  - 被験者をリクルートして、実験に参加してもらう。
  - 主に大学生を対象。
- 研究室実験
  - 研究室の（ある程度）統制した環境下で探索行動を行ってもらう。
- 探索タスク
  - 情報収集タスクを基本に。
  - 自由タスク（固定的な課題やクエリを与えず、テーマを自らの興味、関心に応じて設定してもらい、適宜絞ったり広げたりしてもらう）

# データ収集

- 質問紙
- ウェブブラウザログ
- 画面キャプチャ
- 眼球運動
- 発話プロトコル
- インタビュー
- コンセプトマップ

# 収集データ（1）

## 質問紙

- 事前

- デモグラフィック調査（年齢や主題専攻）
- ウェブやネットメディア経験、知識、スキルについて聞く

- 事後

- 満足度
- 困難さ

- 利用ツール：GoogleDocs等も利用。

# 収集データ（2）

## ウェブブラウザログ

- ブラウザ上でのイベントをそのまま取得
- キーロガータイプ
  - プロキシー系
  - ブラウザネイティブ系
  - OS実行系
    - アドオン系
- キャッシュ機構、ヒストリーデータ、Cookie、POST、AJAX
- 利用ツール：QTHoney,

# 収集データ（3）

## 画面キャプチャ

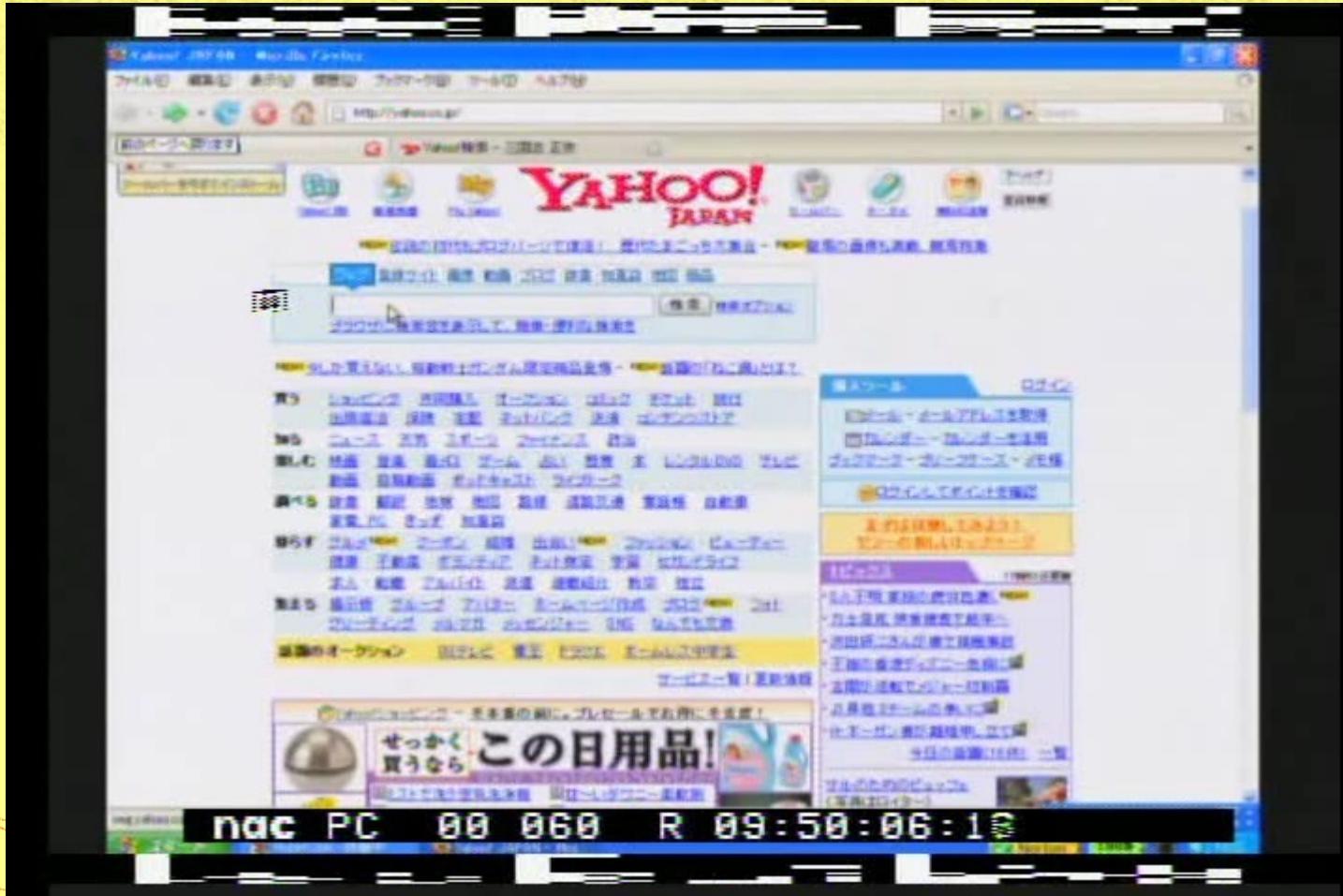
- スクリーンの表示をそのままビデオ映像として取得
- もっとも詳しい情報が得られる。一方で、分析コストはもっとも高い。
- ツールによっては、PC負荷がかかってしまい、探索行動に影響も。
- 利用ツール：HyperCam

# 収集データ（4）

## 眼球運動

- 専門機器カメラと分析ツールによって、眼球の運動を刺激提示モニター上の座標として記録。遷移情報を獲得。
- 被験者の目の特性に応じて取得しやすさや誤差に影響が。キャリブレーションに時間がかかるか、どれくらい探索中の行動に制約がかかるか。
- 分析ツールがウェブ探索等の支援機能を備えているか。
- 利用機器：NAC VOXER, Tobii

# 眼球運動データの例



# 収集データ（5） 発話プロトコル

- 探索中の行動意図を逐一発話してもらう。
- 実験後の分析コストが比較的高い。
  - 書き起こしとタグ付け、アノテーションを行う必要がある。
- 発話を記録するためにマイク等の環境を統制する必要あり。

# 収集データ（6）

## インタビュー

- あらかじめ聞き取りたい内容を統制するかどうか検討しておく。
- 実験後の分析コストが比較的高い。
  - 書き起こしとタグ付け、アノテーションを行う必要がある。複数名でのタグ付け等により信頼性の確保も必要。
- 発話や動作などを記録するためにビデオ、マイク等の環境を統制する必要あり。

# CRESプロジェクトで実施した ユーザ実験事例から紹介

- 実験1. (2007年11月)
  - 収集データ：質問紙、ブラウザ、画面キャプチャ、視線、発話、インタビュー
  - 探索タスク：世界史レポート vs 国内旅行
  - 実験参加者：大学学部生11名
  - 条件：インタビュー中の視線ビデオ提示の有無
- 実験2. (2008年3月)
  - 収集データ：質問紙、ブラウザ、画面キャプチャ、視線、発話、インタビュー
  - 探索タスク：世界史レポート vs 国内旅行
  - 実験参加者：大学院生5名
  - 条件：インタビュー中の視線ビデオ提示の有無
- 実験3. (2009年10月)
  - 収集データ：質問紙、ブラウザ、画面キャプチャ、概念マップ
  - 探索タスク：レポート（メディアの影響、政権交代）
  - 実験参加者：大学生40名（集団実験）
  - 条件：タブ利用の有無
- 実験4. (2010年12月)
  - 収集データ：質問紙、ブラウザ、画面キャプチャ、概念マップ
  - 探索タスク：環境問題記事 vs 国内旅行記事
  - 実験参加者：大学生32名（集団実験）
  - 条件：{発散的, 収束的}探索

# 実験1.と実験2.の概要（タスク）

## ● Web情報探索の課題

- 「世界史」レポートの情報収集
- 国内旅行のための情報収集
  - それぞれ具体的なテーマは実験参加者の興味に応じて決めるとした
- 15分間（各課題遂行の制限時間）

# 実験1.と実験2.の概要 (実験参加者)

- 東京近郊の大学院生と学部生
- 大学院生5名 (男性4/女性1, 平均年齢: 24.6)
  - 図書館情報学専攻 (うち3名が司書資格有り)
  - インターネット利用頻度: 毎日 (4), 週2回以上 (1)
  - サーチエンジン: G (5), G&Y (1)
- 学部生11名 (男性5/女性6, 平均年齢20.0)
  - 専攻は様々: 経済, 工学, 教育, 語学...
  - インターネット利用頻度: 毎日 (7), 週2回以上 (4)
  - サーチエンジン: G (2), Y (5), G&Y (3), MSN (1)

# レポート

たとえば、第2次世界大戦、東インド会社の設立から解散まで、アメリカ合衆国の成り立ちなど

大学の一般教養の授業で、世界史を対象に自分の興味のあるテーマについてのレポートを書く課題が出ました。

テーマは、\_\_\_\_\_にしました。

それでは、レポート作成の事前調査としてインターネットを使って関連資料を集めましょう。調査に使える時間は15分です。役に立つサイトを探しましょう。

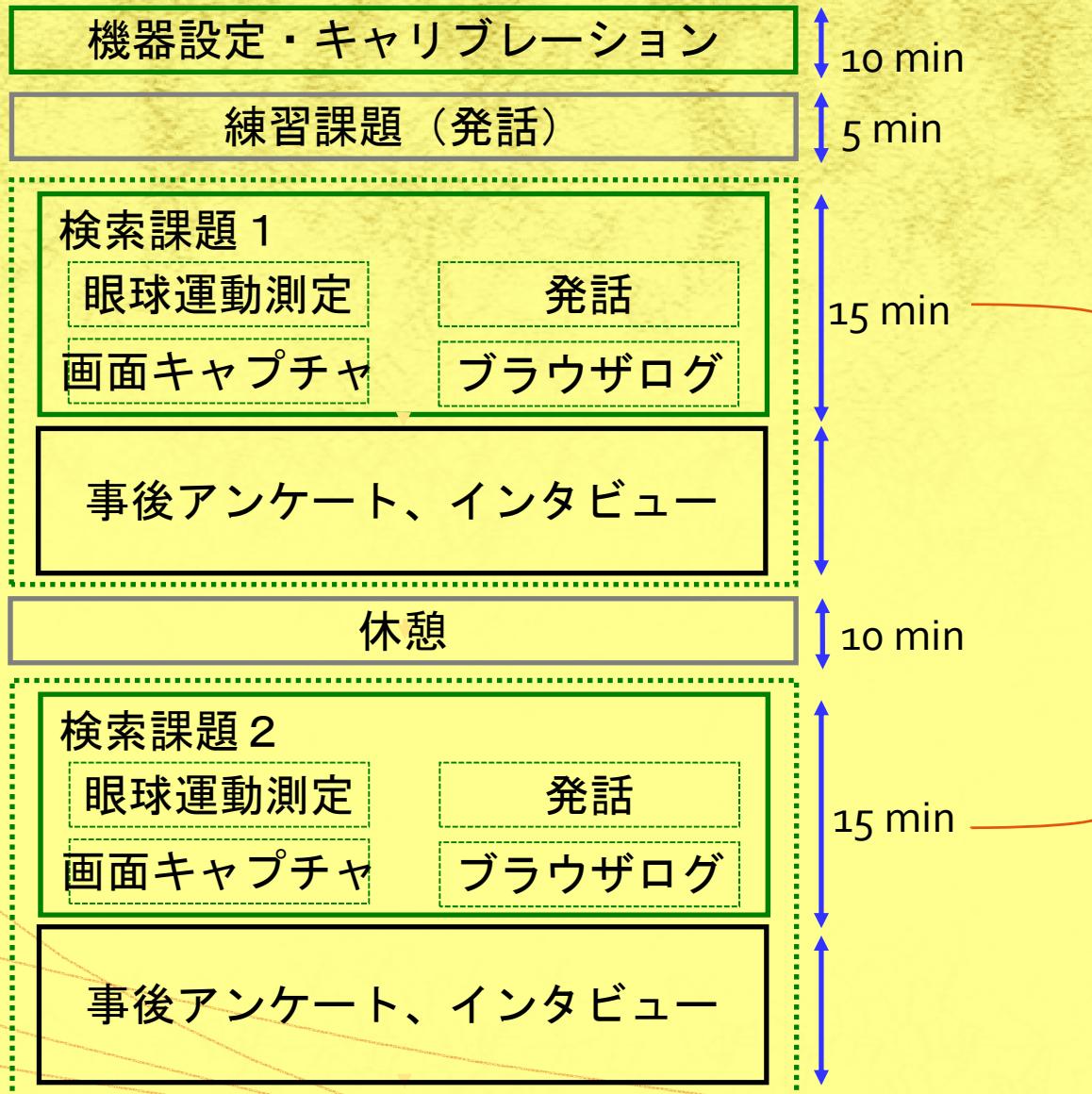
卷之三

# 旅行課題

あなたは、と行く旅行を計画することになりました。

一緒に行く人たちに教えてあげる。たとえば、友達5人と地域への冬休みに沖縄へ、友達2人で春休みに瀬戸内海へ、友達3・4人で冬のスキー場へなど一緒にしましょう。

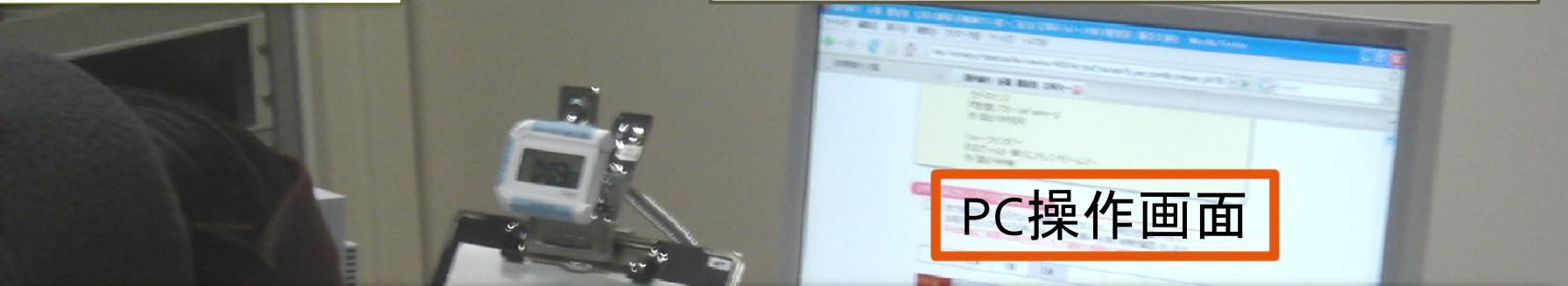
# 実験の流れ



2課題の順序は被験者間で  
カウンターバランス

# 実験環境

19インチ液晶ディスプレイ



画面キャプチャ、ブラウザログ、視線データを計測

被験者

Windows XP

Mozilla Firefox  
(全画面モード)

眼球運動測定装置  
NAC社製 EMR-AT-VOXER

# 実験3.と実験4.の概要 (タスク)

## ● Web情報探索の課題

- それぞれ具体的なテーマは実験参加者の興味に応じて決めることとした
- 15分間（各課題遂行の制限時間）

# 実験3.と実験4.の概要 (実験参加者)

- 東京近郊の大学院生と学部生
- 40名 (男性??/女性??, 平均年齢????)
  - 専攻は様々: 経済, 工学, 教育, 語学...
  - インターネット利用頻度: 毎日 (??), 週2回以上 (??)
  - サーチエンジン: G (??), Y (??), G&Y (??), MSN (??)

## レポート課題

大学の一般教養の授業で、世界史を対象に自分の興味のあるテーマについてのレポートを書く課題が出ました。

テーマは、\_\_\_\_\_にしました。

それでは、レポート作成の事前調査としてインターネットを使って関連資料を集めましょう。調査に使える時間は15分です。役に立つサイトを探しましょう。

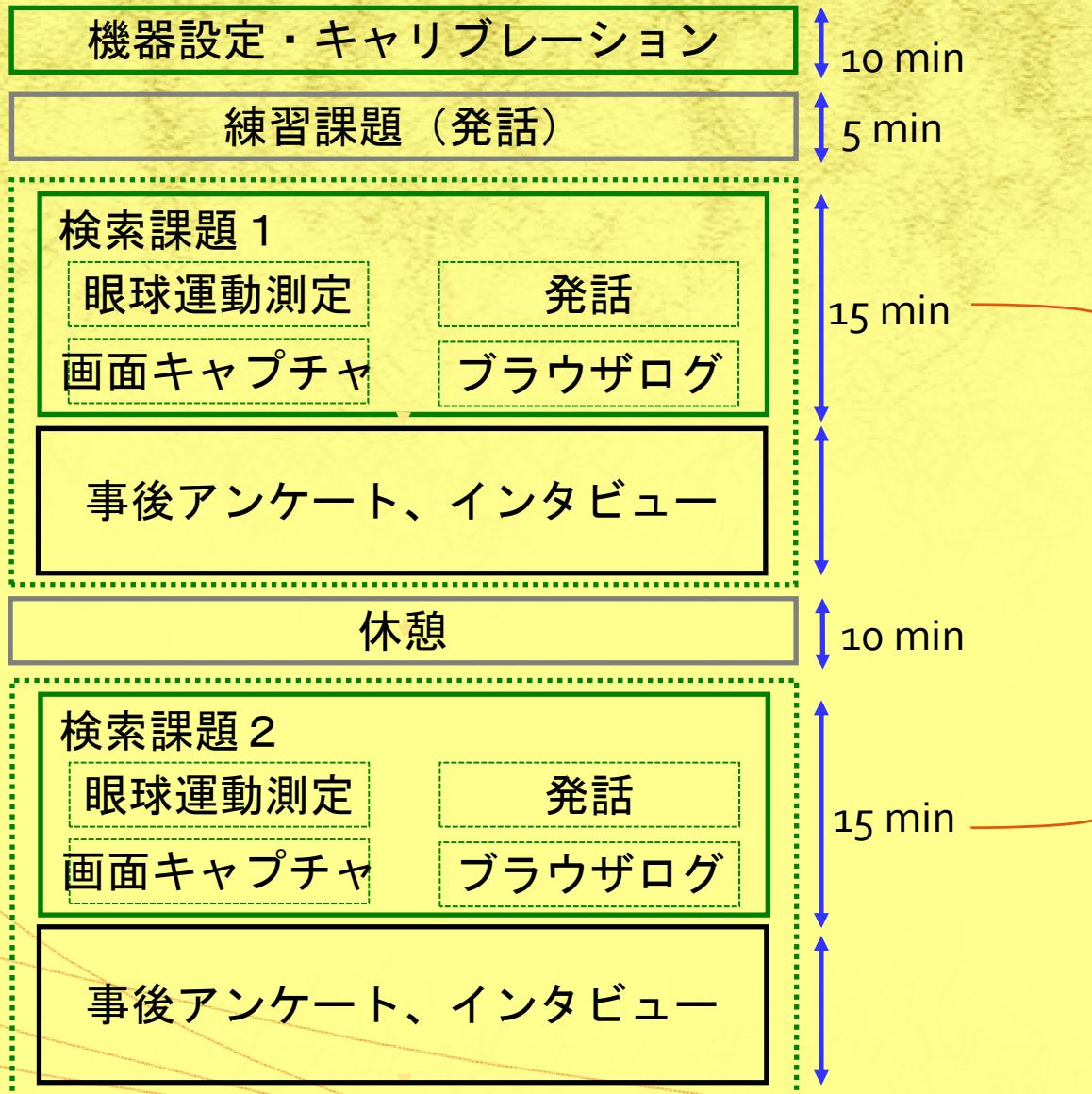
たとえば、第2次世界大戦、東インド会社の設立から解散まで、アメリカ合衆国の成り立ちなど

## 旅行課題

あなたは、\_\_\_\_\_と行く旅行を計画することになりました。  
時期は\_\_\_\_\_で、期間は\_\_\_\_\_、場所は\_\_\_\_\_です。

たとえば、友達5人と冬休みに沖縄へ、友達2人で春休みに瀬戸内海へ、友達3・4人で冬のスキー場へなど

# 実験の流れ



2課題の順序は被験者間で  
カウンターバランス

# 分析手法

- (人手タグ付けデータを元に)
- 探索行動全体
  - ページ分類 : SERP, non-SERP
  - ウェブ行動力テゴリ : 10種類
  - Link Depth
- 視線注視データ
  - Lookzone
  - Scanpath
  - ランキングクリック + 注視

# 分析のため：人手でタグ付け

- 探索行動ログ（ブラウザ、スクリーンキャプチャ）
- 閲覧ページの種類：検索結果一覧ページ、特定ページ
- クリック行動（ランク等）
- ウェブ行動力カテゴリ：10種類
- 視線（アイマーク）位置
  - 0.5秒間隔で検索結果ページ内のLookzoneをコーディング
  - 22カテゴリ
- プロトコル発話
- インタビュー

# 行動の分類

- Web行動力テゴリ: 10種類

- Search: 検索エンジンを使った検索
- Link: リンクのクリック
- Next: 履歴のひとつ先へ進む
- Back: 履歴のひとつ前へ戻る
- Jump: 履歴のひとつ以上前に移動する
- Browse: 別の一覧ページへ移動する
- Submit: フォームなどのボタンをクリックする
- Bookmark: ブックマークに追加する
- Change: ウィンドウやタブを切り替える
- Close: ウィンドウやタブを閉じる

# 閲覧したページの分類

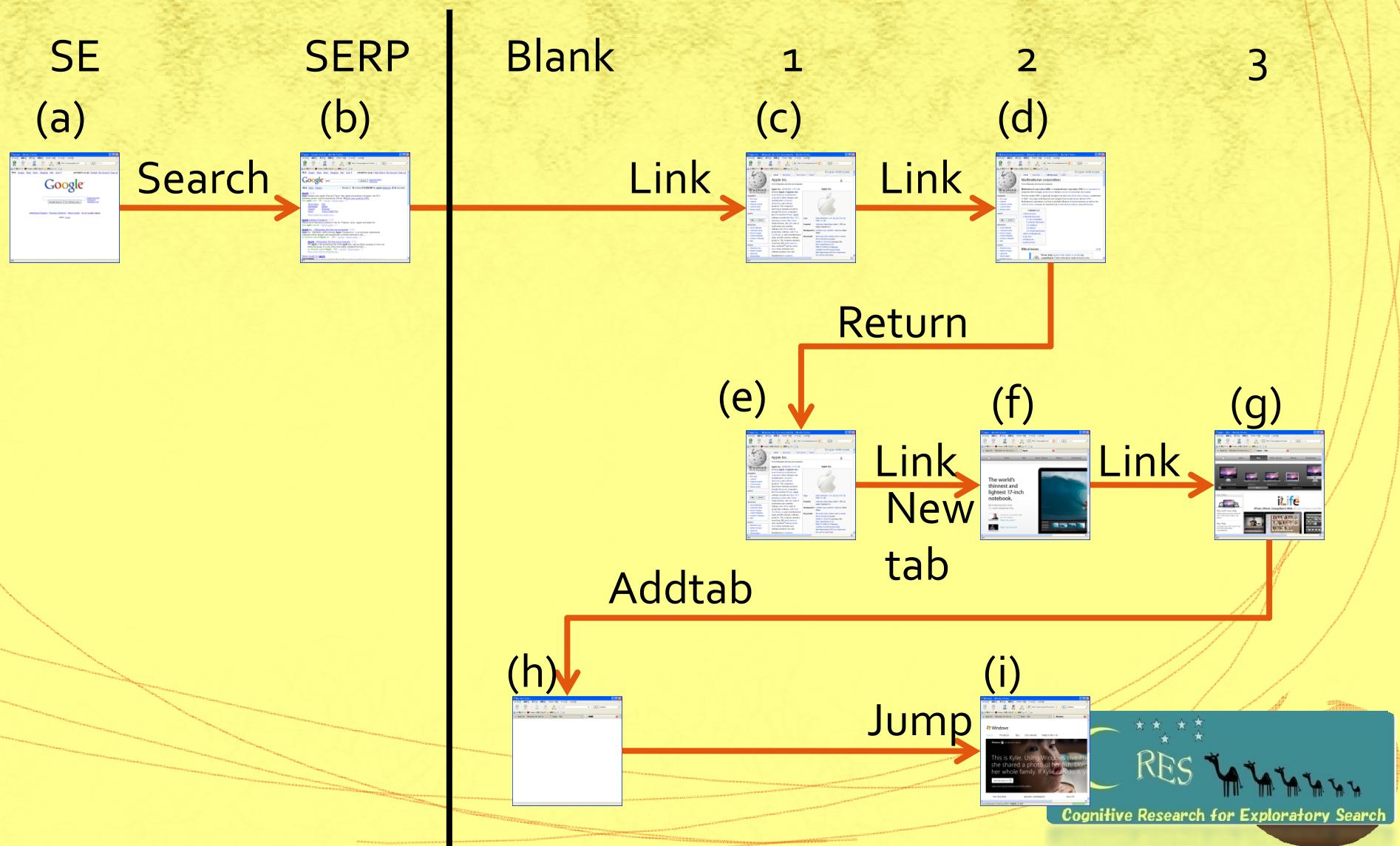
- 一覧ページ  
(SERP)



- 特定ページ  
(non-SERP)



# Link Depthの概念図



# コンセプトマップ

# Lookzoneの分類

1 amason - Google 検索 - Mozilla Firefox

2 ファイル(E) 編集(E) 表示(V) 届(B) ブックマーク(B) ツール(T) ヘルプ(H)

3 ブックマーク(B)

4 リロード(R) 前へ(N) 後へ(P) タブ(T) ショートカット(S) ウィンドウ(W) サイズ調整(+) サイズ調整(-) カラーモード(C) フルスクリーン(F) ヘルプ(H) ハードウェア(H)

5 http://www.google.co.jp/search?q=amason&lr=lang\_ja&ie=utf-8&oe=utf-8&aq=

6 Google

7 検索

8 はじめよう 最新ニュース 上 語数カウント

9 タイトルなし

10 amason

11 検索

12 ログイン

13 amasonに一致する日本語のページ 約 25,600 件中 1 - 10 件目 (0.06 秒)

14 Amazon  
www.amazon.co.jp Amazonはこちら 1500円以上国内配送無料。代引OK

15 もしかして: アマゾン

16 Amazon.co.jp: 通販サイト 本、DVD、CD、ヘルス＆ビューティー ...

17 Photoshop Elements その場で割引き。4月30日までのキャンペーン期間中、対象の『Photoshop Elements』が、その場で1000～4000円割引き。さらに対象商品を購入された方の中から、「Amazonギフト券1万円分」を抽選で20名様にプレゼント。 ...

18 www.amazon.co.jp/-\_113k - キャッシュ - 関連ページ

19 関連検索: 青いケシ 写真, 青海省 青いケシ, 青いケシ ホリデュラ, 青いケシ 植物園, ヒマラヤ 青いケシ 登山家, 皿ガ嶺 青いケシ

20 次へ

21 検索: 次を検索(N) 前を検索(P) すべて強調表示(A) 大文字/小文字を区別

22 完了

# Scanpath

- ランキング部のみに着目
- 各ランク間の遷移を分析
  - (図2の場合)
    - Rank<sub>2</sub> – Rank<sub>1</sub> – Rank<sub>1</sub> – Rank<sub>2</sub> – Rank<sub>3</sub> –
    - = **2-1-1-2-3-4-5-1**

## ● 圧縮Scanpath

$$= \underline{2-1-2-3-4-5-1}$$

## ● Scanpath長

- 各Scanpathに含まれる注視点の数

- 上の例では、

- Scanpath長 = 8

- 圧縮Scanpath長 = 7

図2



# 分析ツールの開発

- COPATT：複数データの統合
- QT Honey：ブラウザログツール
- VizCMAP：コンセプトマップ可視化ツール

# タグ付けツールCOPATTの開発

**スクリーンキャプチャ映像**

**発話音声**

**眼球運動映像**

**タグ入力フォームボタン**

**タグ付け結果**

Position	対象	メタ語ID	メタ語ID	メタ語ID	メタ語ID	メタ語ID	メタ語ID	行動	メタ語ID	メタ語ID	メタ語ID	メタ語ID
13-33.061	特定のページ	1	1	1	1	1	1	focus	1	2	3	4
								submit	Wiki	pedia	フラン	革命
								focus				
								bookmark				

**タイムスタンプ、ページ分類、行動カテゴリ、発話、インタビュー**

20 タイムスタンプ、ページ分類、行動カテゴリ、発話、インタビュー  
21  
22  
23 259.395001 一覧ページ|視線  
24 00:00:00:001 00:00:00:001

Two screenshots of a Japanese Wikipedia page are shown side-by-side. Both screens show the same content: the main page of Wikipedia with a banner at the top stating "28,569 人から寄付を貰いました。" (Received contributions from 28,569 people) and a red button "今すぐ寄付!" (Contribute now!). Below the banner, there's a message about the 11th anniversary of Wikipedia in Japan and the 10th anniversary of the Japanese version. The left screenshot shows a standard browser interface with a toolbar and menu bar. The right screenshot shows a more detailed analysis interface with various toolbars and status bars.

Two screenshots of a video player interface are shown side-by-side. Both are labeled "一時停止" (Paused). The left screenshot shows a video frame with the title "H\$タグ付け\$ysub10\_info\$ysub10\_00002.avi". The right screenshot shows a similar frame with the title "H\$タグ付け\$ysub10\_info\$ym2u00170.avi". Both screens include playback controls like play, pause, and volume, along with a timestamp (00:42 and 00:39 respectively).

対象	行動										注視領域(データは備考1に記録されます)														
	search	close	#####	link	jump	#####	next	change	#####	return	browse	#####	system	submit	#####	focus	bookmark	end	リセット	リセット	▲	▼	◀	▶	リセット
H\$タグ付け\$ysub10_info\$ysub10_00002.avi	42.4453879	H\$タグ付け\$ysub10_info\$ym2u00170.avi	39.8921875																						
対象	対象	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	対象_詳細	行動	行動_詳細	行動_詳細	行動_詳細	行動_詳細	行動_詳細	行動_詳細	行動_詳細	行動_詳細		
13 33.961	特定のページ	1	1	メインページ -	http://ja.wikiped	1	1																		
14 37.6233787																focus									
15 41.7150431																submit	Wikipedia	フランス革命							
16 43.174	特定のページ			フランス革命 -	http://ja.wikiped	1	2																		
17 244.0790136																focus									
18 248.6068231																bookmark									
19 255.6583061																return									
20 256.271	特定のページ			メインページ -	http://ja.wikiped	1	1										return								
21 258.514898																									
22 259.395	一覧ページ	1	22,500,000	Yahoo検索 - ウ	http://search.yah	1											タブ								
23 259.395001	一覧ページ_視線																								
24 260.005001	一覧ページ_操作																								

# QT Honey: ブラウザログツールバー

# VizCMAP: コンセプトマップ可視化ツール

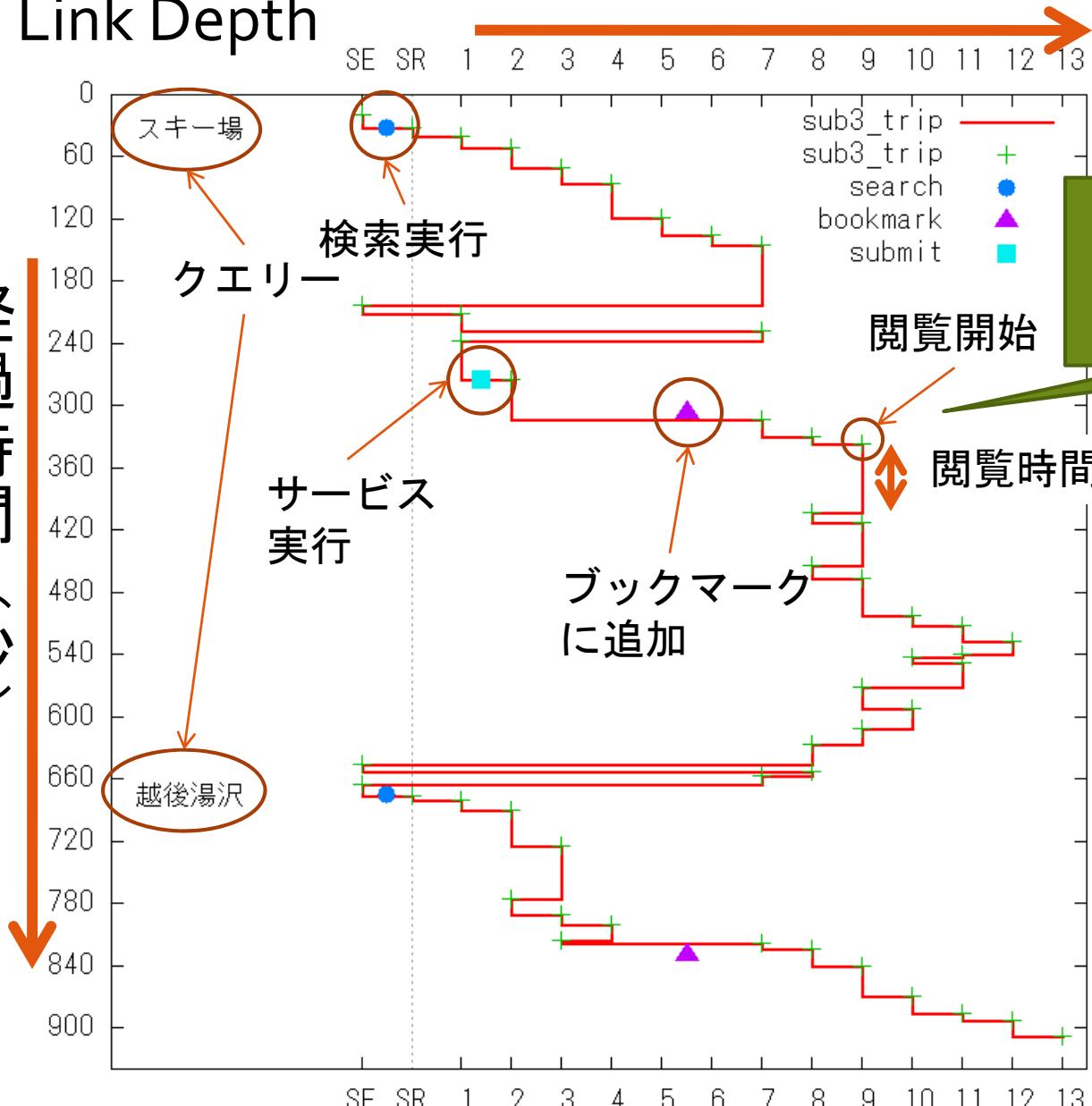
# 探索行動の分析結果

- LinkDepthと行動の可視化
- ウェブ行動力カテゴリ
- コンセプトマップ
- その他

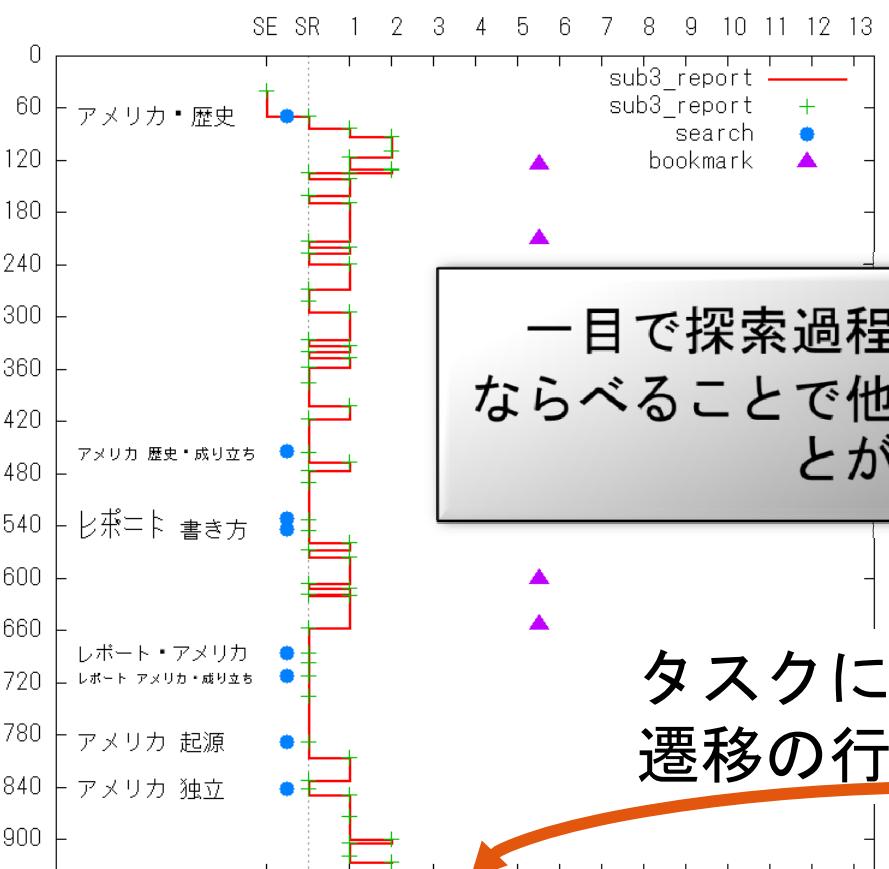
# Link Depthによる探索行動全体の可視化

Link Depth

経過時間（秒）

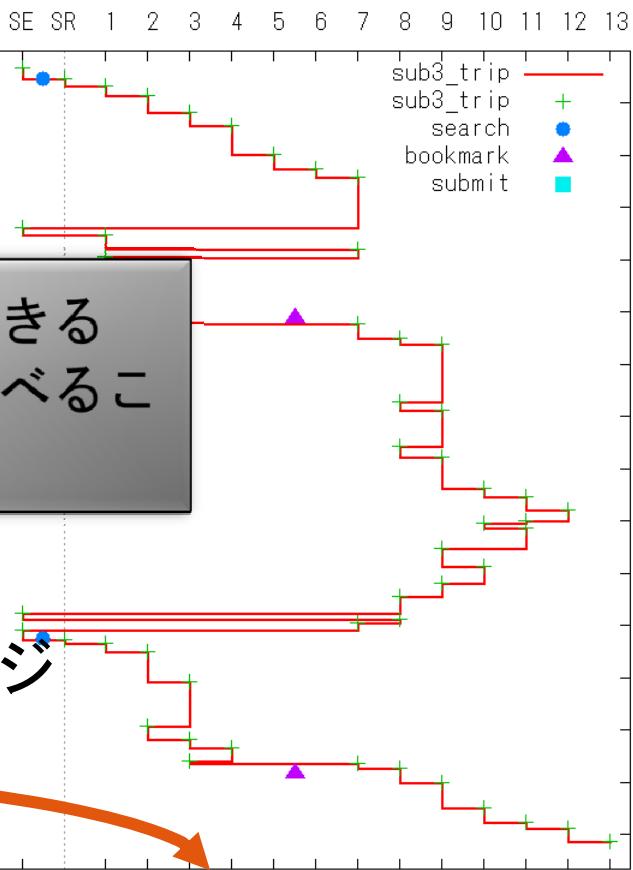


# 可視化例(学部生1人分)



## レポート課題

サーチエンジンの検索結果一覧から直接たどれる結果ばかりを閲覧



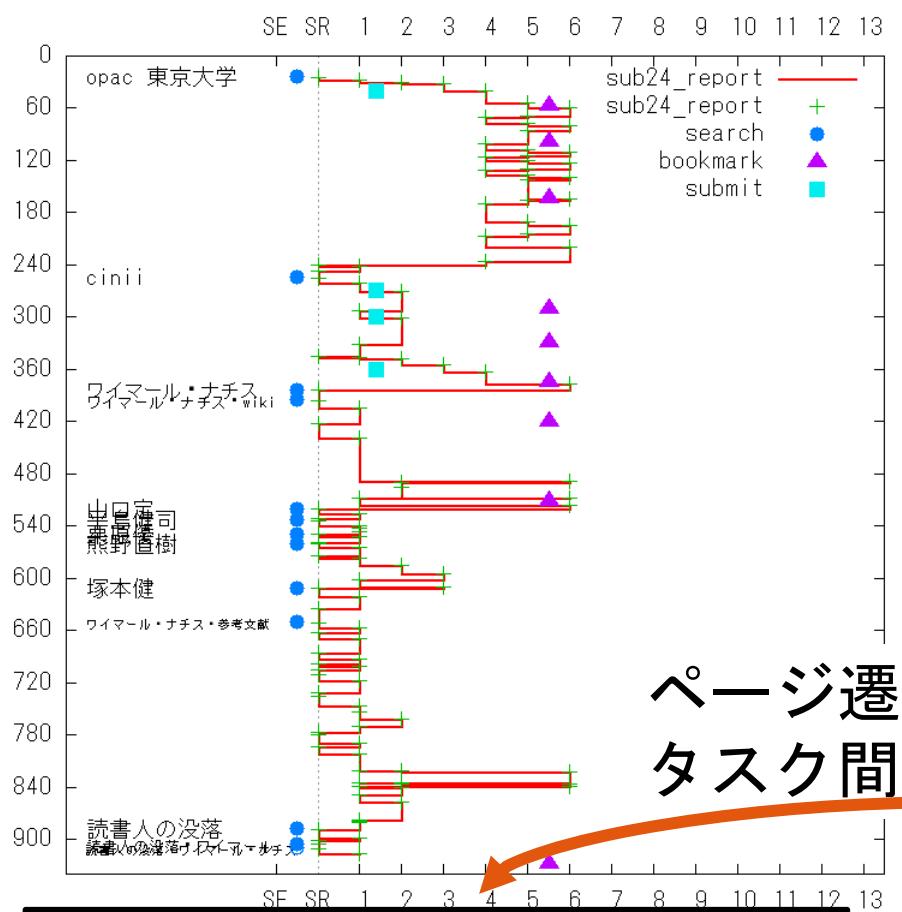
## 旅行課題

サーチエンジンの検索結果一覧ページから離れて深くリンクをたどっている

一目で探索過程全体が概観できる  
ならべることで他のタスクと比べることができる

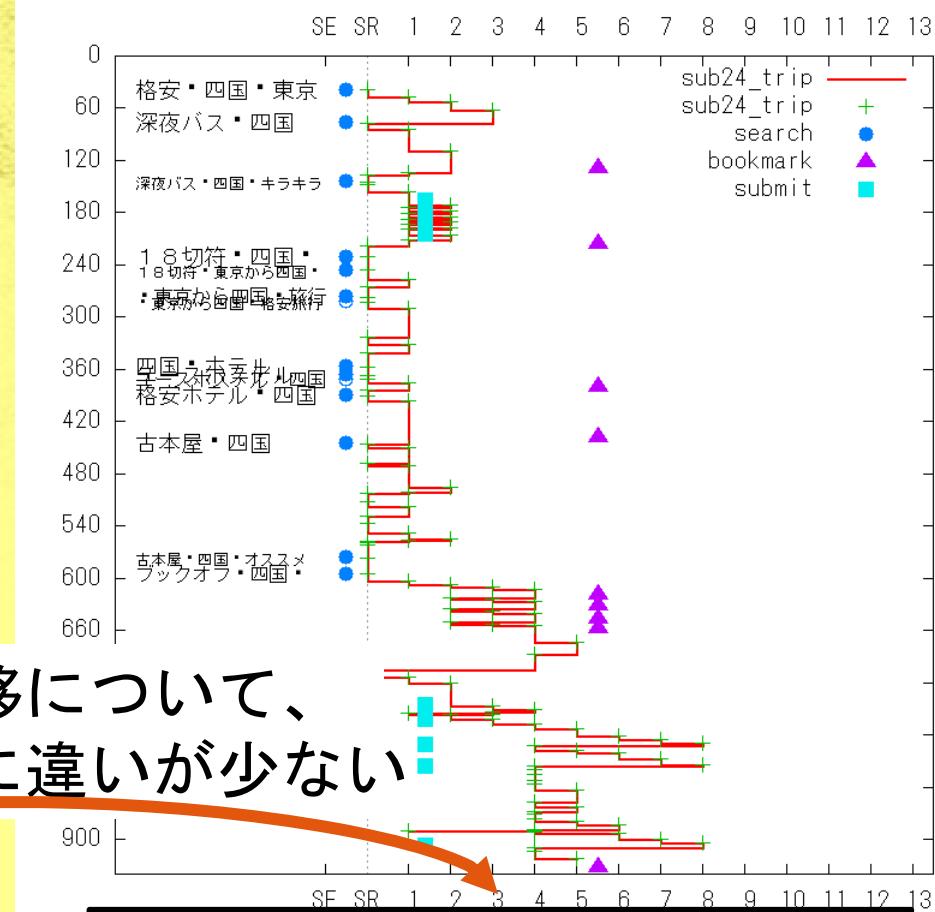
タスクによってページ遷移の行動が異なる

# 可視化例(大学院生1人分)



## レポート課題

サーチエンジンの検索結果一覧から直接たどれるものと深くたどるものと両方ある



ページ遷移について、  
タスク間に違いが少ない

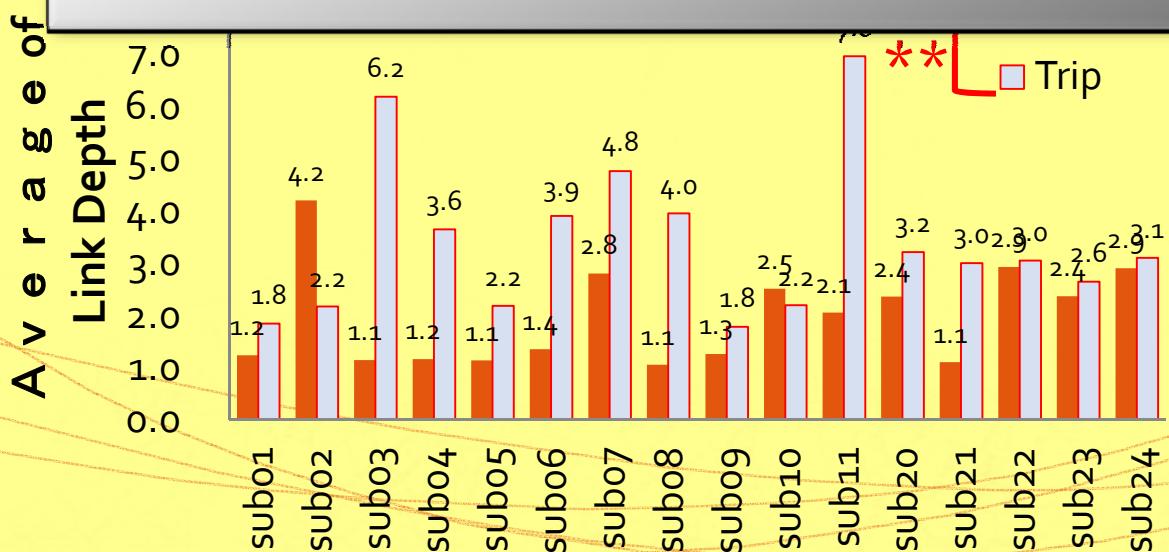
## 旅行課題

サーチエンジンの検索結果一覧から直接たどれるものと深くたどるものと両方ある

# Link Depthの最大値/平均値



旅行課題のほうがレポート課題  
より深いページを閲覧している



# [探索全体]ウェブ行動力カテゴリ

院生はすばやい探索と  
スキャンニング  
学部生は線形的なペー  
ジ遷移

Web カテ ゴリ	院生 (n=11)				学部生 (n=11)			
	平均	SD	平均	SD	平均	SD	平均	SD
Search	3.20	2.55	7.80	5.27	8.00	4.37	6.27	4.92
Link	28.80	7.28	33.20	8.37	19.36	6.26	35.64	8.65
Next	0.80	0.75	0.20	0.40	0.45	0.78	0.91	1.08
Back	10.40	8.11	10.80	7.19	17.45	7.51	22.27	13.80
Jump	2.20	1.72	3.40	2.25	2.64	1.61	2.64	1.92
Browse	0.80	1.17	0.60	1.20	1.82	2.25	0.18	0.57
Submit	7.60	11.29	4.60	4.84	1.27	2.60	3.00	2.80
Bookmark	8.00	1.26	8.00	5.76	4.55	2.06	4.55	2.31
Change	43.60	23.59	28.40	17.85	2.45	5.37	3.55	3.23
Close	4.20	3.54	6.00	8.79	0.36	0.64	2.36	1.77

# 検索結果ページにおけるタスクあたりの各Lookzoneへの平均注視回数

Lookzone	大学院生 (n=5)				学部生 (n=11)				
	レポート		旅行		レポート		旅行		
	平均	SD	平均	SD	平均	SD	平均	SD	
サービスリンク	2.40	2.06	2.20	2.14	17.67	23.44	5.11	9.33	
クエリボックス	5.60	4.36	3.00	4.65	36.89	36.71	12.56	11.93	
検索ボタン	0.00	0.00	0.20	0.40	0.89	1.10	0.67	0.82	
スポンサーリンク	0.00	0.00	11.40	13.99	6.67	7.85	12.44	9.93	
タイトル	41.20	26.80	39.20	40.82	59.67	38.92	42.11	34.19	
スニペット	74.80	42.56	28.40	28.00	91.11	55.59	37.00	32.84	
URL	18.00	9.21	12.40	11.83	40.89	34.27	15.56	11.35	
関連検索	1.20	スニペットやスポンサーリンクを参考に、たどるべきページ内容のチェックや情報の獲得・遭遇が行われている可能性						2.56	4.11
メニュー	1.80							0.11	0.31
ブックマーク	0.00							0.00	0.00
スクロールバー	0.60	0.80	0.00	0.00	0.11	0.31	0.00	0.00	
ツールバー	0.40	0.80	0.40	0.80	1.33	1.63	1.22	1.40	
検索バー	6.40	7.50	4.00	7.04	0.00	0.00	0.00	0.00	
タブ	12.00	14.13	6.00	6.63	8.11	9.81	9.22	17.94	
ステータスバー	0.00	0.00	0.00	0.00	1.78	3.39	0.00	0.00	

# コンセプトマップの分析結果

- ? ?

# 分析結果・まとめ

- Web情報探索実験：探索行動と視線分析
  - {ブラウザログ, 視線注視データ} + タグ付け
  - 2つの要因
    - ユーザ（学部生/院生） vs タスク（レポート/旅行）
  - ユーザ属性：
    - 大学院生の素早い探索過程
    - ブラウザ操作への習熟
  - タスク種別：
    - レポート主題探索の場合の検索結果の読み取りの違い
    - 旅行サイトにおける個別サイト内の深い探索
  - 分析手法の開発
    - ウェブ行動力テゴリ, Link Depth, 注視箇所Lookzone, Scanpath

# 今後の課題 (1)

- 課題はまだまだ残る
  - サンプル数の少なさ
  - より自然な環境への適応
  - もっとも大きな課題：インタビュー・発話プロトコルによる質的分析と量的分析
    - 知識変容タキソノミーの枠組みを用いた分析
      - [三輪, FI'09][Miwa, IR'11]
- 新しい取り組み
  - 別の種類の被験者実験：数種類を試行
    - より人数を増やした集団実験
    - タブブラウズ行動に着目した実験
  - 探索過程の知識変容: コンセプトマップ  
[Egusa, IiX'10][齋藤, JSIK'11]

# 今後の課題 (2)

- ウェブ探索行動の環境, 計測手法の多様化
  - ますます環境・行動パターンは多様に
    - ブラウザが多様に (Firefox 2 → 3 → 4, Chrome, Safari)
    - スマートフォンの普及
  - サーチエンジンの検索対象もさらに多様に
    - ニュース, ブログ/Twitter, 動画, 地図, etc.
- 視線計測技術の研究開発も進展 : Tobii, 安価版ツール等
- 探索ログツールバー QT Honey Toolbar

# CRES:

## Cognitive Research for Exploratory Search

- 本研究はCRES共同研究プロジェクトによるものです
  - 現在の研究成果などは以下のURLからご覧いただけます：
  - <http://cres.jpn.org>
- 本研究成果の一部は以下の研究助成支援を受けています：
  - 国立情報学研究所 公募型共同研究「探索型検索システムに関する認知的研究」
  - 科学研究費補助金 基盤研究(B) No.21300096
  - 科学研究費補助金 特定領域研究（情報爆発） No.18049069

# 参照文献