

NTCIR-9 VisEx の概要

Overview of the VisEx task at NTCIR-9

加藤 恒昭^{1*} 松下 光範² 上保 秀夫³ 神門 典子⁴
Tsuneaki Kato¹ Mitsunori Matsushita² Hideo Joho³ Noriko Kando⁴

¹ 東京大学 ² 関西大学 ³ 筑波大学 ⁴ 国立情報学研究所

¹ The University of Tokyo ² Kansai University ³ University of Tsukuba ⁴ National Institute of Informatics

Abstract: Interactive Visual Exploration (VisEx) is a pilot task at NTCIR-9 for establishing an efficient and effective framework for objectively evaluating interactive and explorative information access environments. It aims to acquire more useful and richer evaluation data based on empirical user studies, by adopting a common framework for the environments and conducting sophisticated experiments. Four teams participated in this task. Although it was harder to understand the results and draw a clear conclusion than expected, we learned much and have made useful progress.

1 はじめに

VisEx は、対話的・探索的な情報アクセスを支援する環境を評価する枠組みを構築することを目的とした取り組みで、情報アクセス技術に関する評価ワークショップ NTCIR-9¹ のパイロットタスクとして実施された。

VisEx では、探索的情報アクセス環境に共通な枠組みを規定し、参加者はこの枠組みの中で動作し、その中心となる核部を提出する。被験者がこの環境を用いて与えられた課題を実施するという実験室実験を行い、情報アクセス環境を評価する。共通の枠組みを設定し、共通の課題を用いることで、評価に影響する要因を限定し、より有益なデータをより多く収集し、評価に結び付けることを目指している。

実験において被験者が行う課題は、与えられたトピックに関して適当と判断される出来事や事実を収集しレポートにまとめるというものである。実験を通じて得られるデータは、この課題の直接の成果物である被験者が作成したレポート、それを作成する過程の情報アクセス環境の利用を記録したログ情報、そして、被験者へのアンケート調査を通じて得られる環境等に関する主観的評価である。これらのデータを個別にそして総合的に解析することで、情報アクセス行為の結果とその時間的側面の関係や客観的評価尺度と主観的評価との関連等に関する知見が得られることを期待している。

本報告では、VisEx のタスク設計の詳細を説明し、参

加システムを簡単に紹介する。加えて、現在までの解析で得られた知見について報告する。実際のところ、得られたデータから意味のある情報を取り出すことは予想以上に難しかった、明らかになった問題点も多い。一方で、それらの問題点を含めて貴重な教訓やデータを得たと考えている。

本報告は NTCIR-9 Workshop meeting での報告 [5] の抜粋である。また、予備実験報告 [4] の内容と重なる部分は割愛している。

2 背景と方針

対話的・探索的情報アクセス環境の評価は難しい。その難しさは、それが支援する情報アクセスが人間の複雑な行為であることに起因する。この行為は直線的なものではなく、試行錯誤や方針の変更が含まれる。それを構成する要素も多様で、その多くは、情報の分析、理解、集約等の創造的な行為である。

このため、対話的・探索的情報アクセス環境の評価では次のアプローチのいずれかがとられる。ひとつは被験者実験を中心とするもので、統制した状況で被験者に情報アクセス環境を用いて課題を実施させ、その過程を観察し達成状況等を計測することで、環境全体を評価する。TREC で実施された Interactive track はその代表例である [1]。課題設計が適切であれば、この方法は実際状況に近いデータを収集できるが、必要となる資源が大きいという問題がある。もうひとつはいわゆるベンチマークテストで、情報アクセス環境の個々の要素を別々に評価する。こちらは比較的安価に実施できるが、その結果が環境の実際の有効性や質を反映しているかに議論の余地が残る。

*連絡先：東京大学大学院総合文化研究科
〒135-8902 東京都目黒区駒場 3-8-1
E-mail: kato@boz.c.u-tokyo.ac.jp

¹<http://research.nii.ac.jp/ntcir/index-ja.html>

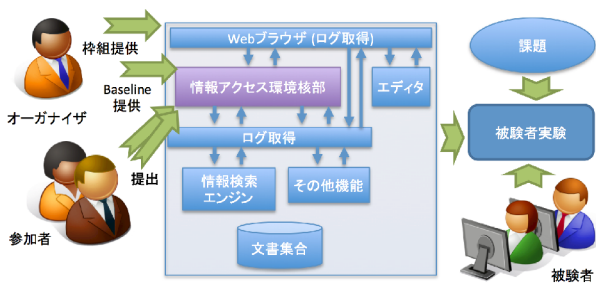


図 1: VisEx の構成

VisEx の最終的な目的は、これら 2 種類の評価を橋渡しし、対話的・探索的情報アクセス環境の評価について、効率的で効果的な方法論を提案することにある。今回の NTCIR-9 における実施では、その第一歩として、被験者実験をより洗練された方法で行って 2 種類の評価の橋渡しを検討できるような様々なデータを収集することを目指した。

このような動機から、VisEx における評価実験は以下の点に配慮している。

- 対話的・探索的情報アクセス行為全体に関する観察を行うこと
- 対話的・探索的情報アクセス行為が自然に生じる課題であること
- 環境の評価にノイズとなって影響するような要因は可能なかぎり排除すること
- 環境全体だけでなく、それを構成する部分要素も観察できること

前 2 者は、被験者実験が対話的・探索的情報アクセス行為全体を観察するものであるために必要であり、後 2 者は、その結果を例えばベンチマークテストによる部分部分の評価と結び付けるために必要である。

3 タスクの設計と実装

3.1 評価の枠組みと情報アクセス環境

図 1 に VisEx における評価の枠組みを示す。VisEx においては、評価される情報アクセス環境は図の中央に示した構成を持つものと仮定される。この環境のうち、核部が参加者によって作成され、提出される。それ以外の枠組部分はオーガナイザより提供され、すべての参加者で共有される。また、オーガナイザは核部のひとつとして、比較のために用いるベースラインも提供する。提出された核部を含んだ情報アクセス環境が共通の課題を用いた被験者実験によって評価される。

情報アクセス環境の枠組では、実際の検索を行う情報検索エンジンや集められた知識の編集や記録に用い

られるエディタ部分を共通化している。更に、すべてが Web ブラウザの下で動作し、利用者はすべてのインタラクションをブラウザ経由で行うとしている。情報アクセス核部は、これらやスニペット生成等のその他の機能を有機的に組み合わせて、情報アクセス環境を構成する。その役割は、利用者の情報ニーズを受け取って情報検索エンジンに送り、その結果得られた情報をわかりやすく、課題の達成を容易にする形で利用者に提示する等して、情報アクセス行為を支援することにある。

このような構成とすることで、(1) 情報アクセス行為全体、つまり、情報を集めるだけでなくそれを知識としてまとめ上げる部分までを観察しつつ、そのような広い行為の観察に伴う揺れをできるだけ少なくする、(2) 情報アクセス環境を用いる利用者の振る舞いに加えて、その中の構成要素間のやりとりについても統一的かつ詳細なデータを取得する、ことを目指している。

実装では、情報検索エンジンはオープンソースの全文検索エンジン Apache Solr²[6] を、Web ブラウザは firefox³ を利用した。スニペット生成等についても Apache Solr のものをそのまま用いている。エディタは、firefox の機能拡張 (アドオン) として開発し、ここにログ取得機能を持たせた。このエディタ及びログ取得機能については、予備実験報告 [4] を参照いただきたい。

オーガナイザが作成したベースラインは、基本的なキーワード検索機能と並べ替え機能を持つものである。予備実験のそれとは異なる実装を行っており、look&feel も若干異なるが、同等の機能を有している。

3.2 課題

実験の課題は、与えられたトピックについての情報を集め、それをレポートにまとめることを被験者は要求するもので、TREC の Interactive track を参考に作成された。以下のふたつの課題を実施した。

イベント収集課題 トピックとして与えられた出来事の特徴、発生日時や発生場所等、を収集し、まとめる。この課題は対話的な高度質問応答と考えることができ、NTCIR-7 で実施された AQLIA [3] の event-list 質問よりトピックを採用している。

- E0 アフリカでなされた恐竜の化石の発掘
- E1 アジアでの航空機墜落事故
- E2 日本で起きた原子力発電所関連の事故
- E3 世界各国で行われた核実験
- E4 N A T O 軍が自ら認めた誤爆

²<http://lucene.apache.org/solr/>

³<http://mozilla.jp/firefox/>

トレンド要約課題 時系列統計情報が関連する社会や経済の状況をトピックとして与えられ、関連する統計量の変化（動向）とその原因や影響を要約する。NTCIR-7で実施された MuST ワークショップ [2] と同じ関心に基づくもので、トピックはそこで作られた時系列統計情報に関するコーパスを参考に作成している

T0 通信関係機器の普及状況

T1 ガソリンを巡る状況

T2 内閣の評価

T3 雇用状況

T4 人口構成

課題のトピックは、あらかじめ、各課題 10 件の候補を参加者に提示して、それを参照して核部の設計や構築を行うことを許している。このような事前の情報開示は普通の評価ワークショップでは行わないが、被験者実験に参加する被験者にとってこれらが新しいことで充分であると判断した。上に示したのはそれらの候補から選ばれ、実際に実験で使われたトピックである。

被験者への教示は、レポートの書式が被験者毎に大きく異なるという予備実験の結果を受け、比較的具体的にいった。まず、全体として、箇条書きにすること、必要な情報のみ抜き出してレポートに含めることを強調している。一方で、トレンド要約課題では、数値のみの列挙となることを避けるために、「ただの数値の羅列ではなく、できるだけ節目となる値や変化を取り出すこと、更にそのような変化の理由やその影響も添えること」を心がけるよう指示している。このため、個々のトピックも、例えば、E1「アジアでの航空機墜落事故について、発生日時、場所、事故状況、航空会社名を含む航空機の種類、死傷者数、原因を調べてください。」、T1「ガソリンを巡る状況の調査として、ドバイ原油価格とレギュラーガソリン価格の変化を調べてください。」のように特徴や統計量を明示して与えている。

文書集合は毎日新聞の 1998-2001 年の記事を利用した。出来事や事態に関する情報が複数記事にまたがり、それらをまとめる等、集約と統合という側面が現れることや、10 年前の記事ということで、予想しなかった事実が見つかったり、新しい解釈が生まれる等があり、被験者の振る舞いに変化がおこることを期待した。

3.3 実験の実施

実験は提出システム（以下、情報アクセス環境核部や情報アクセス環境全体を誤解のない範囲で「システム」と呼ぶ）と課題の組み合わせを単位として行った。一単位の実験では被験者を 5 名とし、以下の順序で実施した。

- 事前アンケートの実施
- 課題の説明およびベースラインを利用した練習トピックについての課題実施の練習（30 分程度）
- 対象システムの説明及びそれを利用した練習トピックについての課題実施の練習（30 分程度）
- 事後アンケートの実施
- 4 つのトピックについての課題実施（50 分）と事後アンケートの実施
- 最終アンケートの実施

すべての被験者はただひとつの単位に参加する。各課題における練習トピック（E0,T0）及び 4 つのトピック（E1-E4, T1-T4）は上に示した通りで、すべての被験者に対してこの順序で実施された。

事前アンケートでは、ウェブ検索やレポート執筆に関する経験の多少等を訊ねている。事後アンケートでは、そのトピックへの親密度、調査結果に満足できたか、難しかったか等を訊ねている。最終アンケートでは使用した対象システムについての評価や感想を求めている。

4 参加チームと提出システム

筑波大学、関西大学、首都大学東京、東京大学の 4 チームが参加し、各 1 システムを提出した。筑波大学、東京大学は両方の課題に参加し、関西大学はトレンド要約課題、首都大学東京はイベント収集課題だけに参加した。筑波大学の Grid システムはキーワードを 2 次元の並びで入力し、ふたつの次元のキーワードの交差する格子に検索結果を配置する。関西大学の KN システムは、統計情報のグラフをインタフェースとして、グラフに描かれた統計情報の関心のある点を指定することで記事の検索が行える。首都大学東京の TM2011 システムは、一度閲覧した記事に既読のマークを付与することができ、新しく検索された記事をそれらと区別できる。東京大学の UTLIS システムは、ベースラインに、記事の出版日と関連する地名による絞り込み検索の機能と指定した記事と類似した記事を探す類似検索機能を付与したものである。⁴

5 結果と分析

被験者実験によって得られるデータは、被験者によって作成されたレポート、課題実施中の被験者の行為やシステムの動作を記録したログ記録、システムやトピックについての被験者へのアンケート結果である。これ

⁴詳細については、*Procs. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies* 及び第 7 回情報編纂研究会予稿集にある各システムに関する論文を参照いただきたい。

表 1: レポート中で参照された記事数

		1			2			3			4		
		min	max	med	min	max	med	min	max	med	min	max	med
E	Baseline	4	12	8	3	10	6	9	14	11	9	18	12
	Grid	5	11	6	4	13	8	6	15	8	7	14	12
	TM2011	5	15	10	4	12	9	8	18	13	11	18	14
	UTLIS	4	12	10	7	12	8	6	11	9	8	13	9
T	Baseline	4	8	6	5	12	9	5	14	7	5	9	8
	Grid	4	8	6	3	16	7	4	13	8	3	13	12
	KN	4	10	10	2	23	10	8	22	9	9	13	10
	UTLIS	5	10	6	4	9	9	6	12	8	6	12	8

らは、評価の対象である情報アクセス環境を用いて行われた課題について、その最終的な成果物、そこに至る動的なプロセス、それを行った主体の印象という異なる側面を表現している。これらを、それぞれに、そして相互に関連づけて解析することで、情報アクセスという複雑な行為を総合的に捉えることができ、その中で情報アクセス環境をの果たしている役割を評価できるのではないかと今回の VisEx での期待である。

しかし、実際には、実験で得られたそれらのデータからその意味が明確な情報を読み取ることは思いの外難しかった。理由としては、被験者間の変動が大きく、加えて、その数も多いとはいえないため、その違いに環境の違いが隠されること、課題の成果物やその過程がどうあるべきかという基準、つまり、何が望ましい姿なのか明らかでないために結果の数量化が困難だったこと、加えて、ログ取得等の実験の仕組みがまだまだ充分洗練されていなかったこと等があげられる。

このような状況であるので、本報告では基本的な分析の結果を報告する。意味のある差が見えていないものも多いが、その一方で様々な観点からの分析が行えている点が重要である。これらが、今後、より深い解析に繋がることが期待される。

5.1 レポート分析

レポートは課題の成果物であり、情報アクセスという行為によって何が得られたかをそこから読み取ることができる。レポートの内容を情報の基本的単位であるナゲット [7] に分解し、その量や質を議論する計画であったが、以下の問題があることが明らかとなった。

あるナゲットをレポートに含めるかは、そのナゲットを見つけ出せたかだけでなく、それを利用者が重要もしくは適合と判断したかに依存する。この判断が利用者によって異なる。加えて、トピックの解釈にも利用者毎に違いがある。これらのことから、評価者において適合ナゲットを定義して、そのナゲットをレポートに含めたか含めなかったかによって、レポートにお

ける適合率や再現率を求めたとしても、これによって見ているものは、環境の違いと利用者の判断の違いの混合物になってしまう。この問題は既に指摘されていたが、比較的判断に揺れが少ないと考えた今回の課題でも同様に生じていることが確認された。

課題への取り組み方の違いもレポートに影響する。例えば、イベント収集課題で、ひとつのイベントについてより多くの情報を得ようとする方向と、より多くのイベントを収集しようとする方向がある。一方の取り組みが、そもそも存在しない情報を探していることにつながり、時間だけ要してレポートに反映される結果が得られない場合もある。このような失敗した取り組みを評価することはレポートの情報からは行えない。⁵

以上の点から、ナゲットによる詳細な分析に先立って、比較的簡単な記事単位の分析を行って、結果の概要を把握することとした。あるレポートにある記事からの情報が含まれているかは、課題の指示として、出典の記事番号をレポート中に記入することを与えているので、機械的な抽出が可能である。適合性の判定は行わず、どの記事がレポート内で参照されているかのみ注目した。

表 1 に、システム、トピック毎に、レポート中で参照された記事数の最小、最大、中央値をまとめた。量的には大きな差や明確な傾向は見られない。また、参照された記事の広がりや違いを見るために、そのシステムを用いた被験者が一人でもある記事を参照していれば、そのシステムによってその記事が検索されたと考え、トピック毎にあるシステムがどのような記事を検索できているかを調べた。表 2 にその結果をまとめると、トピック毎に、検索された記事の総数（タイプ数）と、すべて（4 システム）、3 システム、1 システムにより検索された記事数を示す。後 2 者についてはシステム毎の検索記事数も付記している。ここからシステムの傾向を見ることができる。イベント収集課題における Grid システム、トレンド要約課題における KN システムが、他の 3 システムが検索した記事を検索して

⁵その観察を可能にするのがログ情報である。この点からも情報の総合的な解析が重要となるが、今回はそこに至らなかった。

表 2: システムが検索した記事数

	総数	4 システム	3 システム	1 システム
E1	75	7	8 [5, 7, 6, 6]	46 [13, 7, 16, 10]
E2	108	3	5 [5, 2, 4, 4]	85 [12, 28, 23, 22]
E3	75	10	11 [9, 6, 10, 8]	41 [9, 10, 13, 9]
E4	62	11	12 [9, 5, 12, 10]	28 [5, 5, 5, 13]
T1	52	6	3 [3, 3, 0, 3]	34 [6, 1, 19, 8]
T2	67	3	14 [12, 10, 8, 12]	33 [6, 8, 16, 3]
T3	93	2	14 [9, 9, 12, 12]	56 [14, 10, 21, 11]
T4	51	12	6 [3, 4, 6, 5]	27 [8, 8, 8, 3]

[a, b, c, d] は、イベント収集課題 (E) では、baseline, Grid, TM2011, UTLIS の検索数、トレンド要約課題 (T) では、baseline, Grid, KN, UTLIS の検索数を示す。

いないことが若干多く、検索される記事の傾向がやや異なるようである。

表 2 から、殆どのトピックで、半分以上の記事がひとつのシステムによってのみ検索されており、すべてのシステムによって検索された記事は多くて全体の 20% 程度で、10% を切る場合も少なくないこともわかる。課題設計の面から考えると、このことは正解記事が多すぎることを示唆する。環境の違い、検索方法の違いによって様々な結果が得られるが、それらの結果それぞれにかなりの量の正解が含まれているという状況が想像される。

参照された記事に基づくシステムの特徴に関するもうひとつの指標を表 3 に示す。これは、参照された記事それぞれを次元として、その記事を参照した被験者数を値としてベクトルとし、トピック毎に全システムの平均との余弦係数、ベースラインとの余弦係数を求めたものである。表 2 と同様の傾向が見られることに加え、E4, T4 では全般に参照された記事の広がりがないことが読み取れる。

5.2 ログ分析

利用者のブラウザへの操作については、詳細なログ記録が取得できたが、以下の問題も明らかになった。一部のイベントについては、その単位が細かすぎ、ログの分量が膨大となる。例えば、入力や削除は一字ずつ 1 行となり、スクロールについても、直観的には 1 回であるスクロールが数行に渡って記録される。一方

表 3: 参照記事に基づくシステムの類似性
全システムの平均との類似

		1	2	3	4
E	Baseline	0.76	0.72	0.87	0.88
	Grid	0.78	0.56	0.78	0.82
	TM2011	0.79	0.71	0.86	0.93
	UTLIS	0.83	0.67	0.84	0.81
T	Baseline	0.90	0.84	0.69	0.89
	Grid	0.88	0.75	0.66	0.88
	KN	0.67	0.68	0.77	0.92
	UTLIS	0.85	0.76	0.72	0.94

ベースラインとの類似

		1	2	3	4
E	Grid	0.48	0.28	0.50	0.57
	TM2011	0.44	0.41	0.70	0.75
	UTLIS	0.54	0.28	0.67	0.67
T	Grid	0.76	0.60	0.20	0.70
	KN	0.50	0.35	0.41	0.78
	UTLIS	0.72	0.66	0.38	0.79

で、Ajax による非同期通信や動的なページ書き換えに関しては必要な情報が十分に取得できていない。加えて、データが充分存在していても、第三者がその意味づけを理解することは必ずしも容易ではない。ある行為のその環境での意味が明示されていないためである。例えば、画面へのクリックはその座標と共に記録されるが、その座標でのクリックが記事の選択なのか、次ページへの移行なのかという行為の意味はログ情報の上では自明ではない。

上記のような理由から、参加システムに共通して行ったログ分析は今のところ簡単なものに留まっている。そのひとつとして、知識編集時間の違いを表 4 に示す。ここでの知識編集時間とは、エディタのタブがアクティブになっていた時間を指す。利用者が何らかの有益な情報を発見した後、それをレポートにまとめている時間を近似的に表わしている。全体の時間に対して、どの程度の割合で知識編集が行われていたかを利用者を平均して、システム、課題毎に見ている。イベント収集課題ではベースラインが、トレンド要約課題では KN システムが他に較べてやや少ない値となっている。

5.3 アンケート分析

アンケートで得られた情報の一例として、最終アンケートで調査した、使いやすさ、機能、効率の 3 つの側面で環境を評価した結果を表 5 に示す。以下も含めてここに示した結果は、すべて 7 点のリッカート尺度で取得されたもので、7 点が「多いに同感する」となっている。複数の課題に参加したシステムをみると、ベースラインと UTLIS システムではイベント収集課題での評価がやや高いのに対し、Grid システムではトレンド

表 4: 知識編集時間の割合

		1	2	3	4
E	Baseline	0.42	0.26	0.38	0.31
	Grid	0.45	0.38	0.44	0.43
	TM2011	0.53	0.32	0.40	0.43
	UTLIS	0.44	0.33	0.31	0.36
T	Baseline	0.55	0.60	0.49	0.41
	Grid	0.58	0.56	0.55	0.53
	KN	0.41	0.36	0.31	0.34
	UTLIS	0.56	0.58	0.58	0.41

表 5: システムの評価

		使いやすさ	機能	効率
E	Baseline	5.8	4	4
	Grid	2.8	3.8	3.2
	TM2011	5.6	3.8	4.2
	UTLIS	5.8	4.2	5.2
T	Baseline	4	3.4	3.2
	Grid	5.8	5	5.2
	KN	5.4	4.4	4.6
	UTLIS	5	4	4.8

要約課題での評価が高く、システムの設計と課題との相性が示唆されている。

最終アンケートでは課題についての印象も尋ねている。課題設定について言えば、難しいかという評価で 4.3、複雑であるかでは 3.3、時間を要するかでは 5.0 の平均を得ており、比較的単純であるが時間がかかる課題と評価され、この点でも適合情報が多すぎるトピックであったことが伺える。

個々のトピックの終了後に行ったアンケートでは、トピックへの親密度、その調査への満足度、その調査を困難と感じたかを訊ねている。幾つかの項目について求めたその間のピアソンの積率相関係数を表 6 に示す。トピックへの親密度（馴染み）と調査の難しさにはほとんど相関がなく、必要な情報の量の多少等が影響していることがわかる。また、調査への満足は情報を十分に集められたという感触に強く裏付けられていることも観察できる。

表 6: トピックに関する印象の相関

	調査結果に満足している	難しかった
難しかった	-0.39	-
トピックに馴染みがある	0.18	0.01
時間は充分だった	0.50	-0.23
十分に情報が集められた	0.75	-0.38
利用できる記事が豊富だった	0.56	-0.31

6 おわりに

対話的探索的情報アクセスを支援する環境についての評価方法を確立するために、VisEx を実施し、4 チームの参加を得た。得られたデータから明確な結論を得ることは思いの外、難しかったが、学ぶことは多かった。タスクの設計においては、もう少し複雑な課題とすること、実験デザインとして、被験者間のばらつきを吸収する仕組みを導入することが大事だろう。ログの取得方法も提出されたシステムに応じて考えていく必要がある。基本的な枠組みについては、情報アクセスという複雑な行為の様々な側面についてのデータが統一的に取得できる点があることが確認された。今回の実験で得られたデータについても今後詳細な検討を進める必要がある。結論として、今回の VisEx によって、我々は貴重な一歩を踏み出したといえる。

謝辞

本研究の一部は基盤研究 (B) 「視覚情報を活用した対話的情報アクセスのための情報編纂研究基盤の構築」の一環として行われ、科学研究費補助金により支援されています。国立情報学研究所との共同研究として進められている部分もあります。ご支援をここに深く感謝します。

参考文献

- [1] S.T. Dumais and N.J. Belkin: The TREC Interactive Tracks: Putting the User into Search. in E.M. Voorhees and D.K. Harman ed. TREC Experiment and Evaluation in Information Retrieval, pp. 123-152, The MIT Press, 2005.
- [2] T. Kato and M. Matsushita: Overview of MuST at the NTCIR-7 Workshop - Challenges to Multi-modal Summarization for Trend Information. In *Procs. of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, pp. 475-488, 2008.
- [3] T. Mitamura, E. Nyberg, et al.: Overview of the NTCIR-7 ACLIA Tasks: Advanced Cross-Lingual Information Access. In *Procs. of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, pp. 21-25, 2008.
- [4] T. Kato, M. Matsushita and H. Joho: Overview of the VisEx task at NTCIR-9/ In *Procs. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, 2011.
- [5] 加藤恒昭, 松下光範, 上保秀夫: VisEx 予備実験報告. 第 5 回情報編纂研究会, 2011.
- [6] 関口宏司, 大谷純他: Apache Solr 入門. 技術評論社, 2010.
- [7] E.M. Voorhees: Question Answering in TREC. in E.M. Voorhees and D.K. Harman ed. TREC Experiment and Evaluation in Information Retrieval, pp. 233-257, The MIT Press, 2005.