

## レlevanceをめぐる一考察 - NTCIR の背景として -

野末 俊比古      神門 典子

学術情報センター研究開発部

{nozue,kando}@rd.nacsis.ac.jp

我々は現在、情報検索システム評価用のテストコレクション(NTCIR)の開発を進めている。テストコレクションにおいては、正解文書セットを作成する際に、文書が検索要求に対して正解かどうかを判定する「レlevance判定」の作業を必要とする。レlevance判定は非常に重要、かつ非常に困難な作業であり、その質がテストコレクションの質を決めるといってもよい。レlevanceについては、情報検索研究の中心的テーマとして、これまでに様々な視点から多くの研究がなされてきた。本稿では、テストコレクションへの適用という観点から、レlevanceに関する先行研究を整理し、NTCIRにおけるレlevance判定の考え方について述べる。

### Primary Considerations in the Concept of Relevance: Relevance Judgement of NTCIR

Toshihiko NOZUE      Noriko KANDO

Research and Development Department  
National Center for Information Science Systems (NACSIS)

We are now developing a test collection for information retrieval systems, in which relevance judgement is necessary. Though relevance is an important topic in IR research, there is little agreement on "what is relevance". This paper examines the concept of the relevance from the view point of the test collection.

#### 1 はじめに : テストコレクション

我々は、日本語情報検索システム評価用のテストコレクション (NACSIS Test Collection for Information Retrieval; NTCIR) の開発を進めている[1][2][3][4]。現在、コンペティション形式のワークショップを通して、NTCIR-1 を構築中である。本稿では、テストコレクションや情報検索システムの評価において根本的な概念であるレlevanceについて論じる。

##### 1.1 テストコレクションとは

テストコレクションとは、情報検索システム評価のための検索実験用のセットのことで、(1) データベース、(2) 検索課題、(3) 各検索課題に対する

データベース中の正解文書の網羅的リスト、からなる。情報検索のシステムや手法の評価に用いられる再現率を計算するには、データベースに含まれる正解文書の総数が必要である。しかし、現実の情報検索ではこれは常に未知であるため、再現率は算出できない。これに対し、テストコレクションでは、予め網羅的な正解文書リストがつくられているので、再現率の計算が可能になる。近年、新しい情報検索のシステムや手法の妥当性や有効性を実証するには、テストコレクションを用いて検索実験を行うことが必須になってきている。

##### 1.2 テストコレクションにおけるレlevance判定

テストコレクションでは、正解文書リストを作成するために、ある文書が検索課題に対して「正

解」であるかどうかという判定、すなわち、レlevance判定を行う必要がある。テストコレクションを用いた検索実験による評価は、この「レlevance」という概念によって成立しているといえる。

レlevance判定は、容易な作業ではない。テストコレクションにおけるレlevance判定は「客観的」に行う必要があるため、一般には、言語（文章）で表現された検索課題（トピック）に対して、文書（から読みとれるトピック）が「適合」といえるかどうかを判定者が判断する。しかし、複数の判定者が同一の文書に対して行なった判定の一致度は3割に過ぎない、という報告があることから判るように[5]、レlevance判定を厳密に客観的に行うことは難しく、判定者の主観的な判断の影響を受けざるを得ない。さらに最近では、より大規模なテストコレクションが要求されるようになっており、対話型検索システムが主流になっており、従来のバッチ型を想定していたやり方では対応が難しくなっていることなどから、レlevance判定はいっそう難しい作業となってきた。

### 1.3 研究の目的

このような状況を考えると、レlevance判定の背景や基盤となる問題として、そもそもレlevanceとは何か、という議論を整理しておくことは必要かつ有効である。本稿では、テストコレクションにおけるレlevance判定のための基礎的考察として、レlevance概念をめぐるこれまでの議論を整理、検討する。そして、NTCIR-1における考え方がそのなかにもどう位置づけられるのかを示す。

レlevanceに関する研究は、哲学や心理学などを含む様々な学問分野で行われてきたが、本稿では、情報検索を扱う学問分野、すなわち、広い意味での情報学に焦点を当てる。レlevanceを中心的なテーマとして取りあげており、レビュー的な観点を含んだ論文から主要なものを分析した。

## 2 レlevanceをめぐる二つの立場

### 2.1 レlevanceの辞書的な定義

レlevanceは現在、一般には次のように定義さ

れている[6]。なお、‘relevance’の日本語訳としては「適合性」もあるが、比較的狭い意味で使われる傾向もあるので、本稿では、より意味的に広がりのある relevance 本来のニュアンスを出すため「レlevance」と表現する。

#### 適合性 relevance

情報検索において、ある情報要求あるいは検索質問に応じて検索を行なったとき、検索された情報が、どの程度、検索者の情報への欲求を解消したり、質問（疑問）に解答を与えることができているかの度合い。情報検索の主要概念にもかかわらず、その捉え方についてはさまざまな解釈がある。大きく分けると、検索質問と情報との関係と、情報要求と情報との関係の二つの立場から捉えることができる。前者は、質問としての言語表現と情報との関係に着目し、曖昧さをある程度排除することで検索評価において一般的に使われてきた。後者は、情報要求を持つ利用者の観点から適合性を見るもので、情報検索が利用者の要求する情報を提供するものであるならば、こちらの方が適合性のより本質的な解釈といえる。

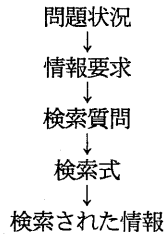
この定義にあるように、レlevanceは情報検索研究における“主要概念”であり、これまでに多くの議論がなされてきたにもかかわらず、現在でも“さまざまな解釈”が存在しており、(辞書的な定義を越えた) 確定的な定義や解釈が一般に形成されているといえる状況にはない。上記の定義にある“二つの立場”から議論を始めることとしたい。

### 2.2 情報探索過程と二つのレlevance

情報検索は、情報探索過程という大きな流れのなかで考えることができる。情報探索過程については一般に、図1のようなモデルが用いられる。問題状況とは利用者が直面している問題であり、これを解決するために生じるのが情報要求である。すなわち、情報要求とは、利用者一人一人が置かれている状況や関わっているタスクなどの途中で、「知識の変則状態 (ASK)」[7]が生じ、それを解決しようとする人間の自然な要求であり、状況やタスクと密接な関係にある。情報要求を言語によって利用者の外部に表現、すなわち客体化したものが検索質問である。さらに、これを検索システムが理解できるかたちにしてシステムに投入した

ものが検索式である。

図1 情報探索過程のモデル



このモデルによって、先の“二つの立場”を捉えれば、前者は検索質問（あるいは検索式）と検索された情報との関係、後者は情報要求と検索された情報の関係をもってレレバンスを捉えている。前者を「システム志向のレレバンス」、後者を「利用者志向のレレバンス」と呼ぶことにする。

レレバンス研究はいろいろな視点からなされてきたが、その多くは、この二つの立場のいずれかにおいて捉えることができる。それぞれの特徴をまとめると、表1のようになる。

表1 二つの立場のレレバンス

システム志向	利用者志向
トピック依存	状況依存
客観的	主観的
測定可能	測定困難
「レレバンス」	「パーティネンス」

### 2.3 これまでのレレバンス研究

二つの立場を軸に、レレバンスをめぐるこれまでの研究の流れを振り返っておく。レレバンスは当初、情報検索システムの評価に必要な概念として登場し、数値的計測可能性の高いシステム志向のものとして取りあげられていた。しかし、トピックの一致度だけで判定する方法に対し、それだけで検索された情報のレレバンスが判るのかを疑問視する声があがった。

利用者を切り離れたレレバンスの考え方への批判として、利用者に関わるタスクや状況まで考え合わせた利用者志向のレレバンスの必要性や重要

性が主張されるようになった。検索システムの利用者がレレバンスを判定する際に、どのように判定しているかを実際に調査した結果からは、検索の目的、費やす時間、著者など、トピックではなく、状況やタスクに関わる要素が大きく影響しており、同一の人間でも状況によって判定が異なることが示された[8]。

しかし、利用者志向のレレバンスは個別的なものであり、ときどきで判定が異なるものであり、一般化が難しく、大量には扱いづらいため、検索システムの性能評価に導入するのは困難なことが多い。現実には、あらゆる種類の検索質問がありうるが、これに耐えられるロバストなシステムの評価には統計的に有意さが求められる。利用者志向の立場から批判は高まってきたものの、特定の利用者を想定せずにトピックを取り出し、実験的環境を成立させるシステム志向の方法に代わる簡易な方法を登場させるには至っていないため、依然として、システム志向のレレバンスの考え方は有効であり続けてきた。

### 3 レレバンスの新しい捉え方

二つの立場の違いを前提や目的の違い、別個のものとして扱うこともできる。しかしながら、システム志向の立場に関しては、「本当に利用者の状況と完全に切り離して、トピックだけでレレバンス判定ができるのか」、逆に利用者志向の立場に関しては、「状況がレレバンス判定に影響を及ぼすといってもトピックを欠かすわけにはいかないのではないか」といった疑問を呈することができる。これまで、必要以上に対立してきた感のある二つの立場は、必ずしも排他的なものではない、と考えられないだろうか。

そのような、二つの立場を越えた観点からレレバンスを捉える作業は、これまで余りなされてこなかった。しかし、最近、俯瞰的な視点からの研究が見られるようになってきた。ここでは、そうしたなかから、Saracevic と Mizzaro の研究を中心に取りあげ、検討してみたい。

表2 Saracevicによるレレバンスの分類 (文献[9]より作成)

レレバンス	関係要素	判断基準
システムレレバンス アルゴリズムレレバンス	「検索式」と「システムのファイル中の情報 オブジェクト (テキスト)」	検索性能
トピックレレバンス 主題レレバンス	「検索式に表された主題/トピック」と「検索 された/システムファイル中の/現実のテ キストでカバーされたトピック/主題」	アバウトネス
認知レレバンス 適切性	「利用者の知識や認知的な情報要求の状態」 と「検索された/システムファイル中の/現 実のテキスト」	認知的調和、報知性、新奇性、 情報の質など
状況レレバンス 効用	「状況、タスク、問題」と「検索された/シ ステムファイル中の/現実のテキスト」	意志決定における有用性、問 題解決における情報の適当 性、不確実性の軽減など
動機レレバンス 感情レレバンス	「利用者の意志・目的・動機」と「検索された /システムファイル中の/現実のテキスト」	満足、成功、達成など

### 3.1 Saravevicによるレレバンスの分類

Saracevic は、最近の論考[9]のなかで、レレバンスを考える枠組みと、レレバンスの定義づけの議論とを分けて論じている。Saracevic によれば、レレバンスを考える枠組みとして、これまで、システムの枠組み、コミュニケーションの枠組み、状況の枠組み、心理学の枠組みの四つがあった。そして、それぞれの枠組みの利点と欠点を指摘した上で、新しい枠組みとして、システムとの対話の枠組みを提案している。対話という枠組みで捉えるならば、現実のレレバンスは、利用者とシステムとの相互作用的なやりとりのなかで変化していくものであり、そのときどきでレレバンスも様々に存在することを示唆している。

さらに、Saracevic は、これまでのレレバンスの定義をめぐる議論から、これらを五つに分類できるとした(表2)。この分類を情報探索過程のモデル(図1)を使って解釈すれば、システムレレバンスは検索式と情報、トピックレレバンスは検索質問と情報、認知レレバンスは情報要求と情報、状況レレバンスは問題状況と情報との関係に着目したものだといえよう。動機レレバンスは情報探索過程を全体的にカバーする、より一般的、抽象的なもので、情報と「検索式から問題状況を含む全体、あるいはそれを越えたもの」との関係に着目したといえよう。よって、従来の二つの立場(表

1)で捉え直せば、システムレレバンスから状況レレバンス(動機レレバンス)に向かうにしたがって、システム志向の特徴が弱くなり、利用者志向の特徴が強くなっているといえる。

このように、レレバンスは、対話という枠組みから、時間の経過とともに変化していくものであり、そのときどきで(五つに分類が可能な)多様なレレバンスが動的に存在し、全体として一つの体系をなすような複合体として捉えられることができる。そこでは、これまでの二つの立場は、排他的なものではなく、いずれの特徴が強いか、という相対的なものとして反映されているに過ぎない。

### 3.2 Mizzaroによるレレバンスの解釈

Mizzaro は、これまでのレレバンスをめぐる議論を包括的な立場からまとめ、レレバンスを四つの次元の関係として解釈する枠組みを示している(表3)[10]。ここで、第一の次元を文献(情報)の次元( $I_n$ )、以下、利用者の次元( $U_n$ )、ドメインの次元( $D_n$ )、時間の次元( $T_n$ )と呼ぶことにする。

これまで、提唱されてきた様々なレレバンスの定義は、この枠組みでは「どの時点(時間の次元)で、どういう視点(ドメインの次元)において、文献(情報)の次元と利用者の次元のどの要素とどの要素の関係を描いているか」によって表現される。四つの次元をそれぞれ変数と見たとき、四つの変数の組合せで表現される変数がレレバンス

表3 Mizzaroによるレレバンスを構成する四つの次元（文献[10]より作成）

第一の次元	I <sub>1</sub> 文献表象要素	タイトル、キーワードリスト、著者名、書誌データ、要旨、抽出文など
	I <sub>2</sub> 文献	物理的な意味での文献
	I <sub>3</sub> 情報	文献を読んで得られる情報
第二の次元	U <sub>4</sub> 問題状況	解決に情報を必要とするような、人間が直面している問題
	U <sub>3</sub> 情報要求	利用者の内面における問題状況の表象物
	U <sub>2</sub> 検索質問	情報要求の“人間”言語による表象物
	U <sub>1</sub> 検索式	情報要求の“システム”言語による表象物
第三の次元	D <sub>1</sub> トピック	利用者が関心を持っている主題領域
	D <sub>2</sub> タスク	利用者が検索された文献を用いて実行する活動
	D <sub>3</sub> 文脈	トピックやタスクに属さない、検索および検索結果の評価の方法に影響するすべて
第四の次元	T <sub>n</sub> 時間	問題の発生から解決までのある時点

※ D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>は、排他的ではない。

といってもよいだろう（レレバンス判定はその値を決定する作業といえる）。

Mizzaro もまた、レレバンスは動的に変化するものであると考え、時間の次元を設定している。さらに、情報検索過程に対応して、利用者の次元の四つの要素を挙げている。文献（情報）の次元として三つの要素を挙げているが、これは、情報検索過程における「検索された情報」を、検索結果（二次情報）、検索結果をもとに入手する文献等（一次情報）、文献等を読んで得られる情報（の内容）、という三つに分けて考えたものといえる。

### 3.3 複合体としてのレレバンス

Saracevic と Mizzaro の議論はいずれも、表現の仕方は異なるものの、レレバンスは単独のものではなく、様々な要因が関わっている、複数の（種類の）複合体であり、それをメタな視点から捉えることの重要性を主張しているものと理解したい。Saracevic の分類によるレレバンスを Mizzaro の枠組みによって解釈すれば、概ね、

- ・システムレレバンス：I<sub>1</sub>, D<sub>1</sub>, U<sub>1</sub>
- ・トピックレレバンス：I<sub>1-2</sub>, D<sub>1</sub>, U<sub>2</sub>
- ・認知レレバンス：I<sub>1-3</sub>, D<sub>1-2</sub>, U<sub>3</sub>
- ・状況レレバンス：I<sub>1-3</sub>, D<sub>2-3</sub>, U<sub>4</sub>
- ・動機レレバンス：I<sub>1-3</sub>, D<sub>3</sub>, U<sub>4</sub>

のようになろう。これを図示すれば、図2のようになる（I, Tは捨象）。グラフの右上ほど利用者志

向の特徴が強く、左下ほどシステム志向の特徴が強い（動機レレバンスは、「情報」と「問題状況」の外側にある様々な要因」との関係と捉えた）。

## 4 NTCIRにおけるレレバンス判定

情報検索システムの評価という観点からは、Saracevic や Mizzaro のような俯瞰的な視点から捉えたレレバンスが、テストコレクションにおけるレレバンス「判定」にどのように影響しうのか、という点が重要となる。以下、NTCIR-1におけるレレバンス判定の考え方を述べる。

### 4.1 NTCIR-1の概要

我々は現在、コンペティション形式のワークショップを通して、NTCIR-1を構築中である。NTCIR-1の意義は次のような点にある。

- ・大規模な日本語テストコレクション
- ・文書のジャンルの多様化（学術的文献）
- ・多言語・言語横断検索への対応

NTCIR-1は、

- ・文書データベース
- ・検索課題
- ・検索課題に対する正解文書のリスト

の三つからなっている。文書データベースにはNACSIS-IRの「学会発表データベース」のデータの一部を使用している。文書は、国内の65学

会の大会や研究会における発表論文の抄録約 33 万件であり、半数以上は日本語・英語の対訳となっている。また、一部については語構成要素も考慮した詳細な品詞タグを付与した。正解文書リストは、データベース中の文書のうち、各検索課題に対して正解となるものの網羅的リストであり、人手によって正解判定を行なっている。

検索課題は、利用者の検索質問を一定の形式で記述したもので、実際に研究者から収集した。ワークショップでは、訓練用 30 件、評価用 53 件の課題を用いているが、最終的には 300 件になる予定である。検索課題は、次のような構成になっている。

- ・タイトル：検索要求を単語・フレーズで短く表現したもの
- ・検索要求：「○○について論じた論文が欲しい」など、検索質問を自然言語で書いたもの
- ・検索要求説明：文献が満たす条件・要素、検索の目的、用語の解説など、検索要求の補助的な説明
- ・概念：検索要求に関係するキーワード
- ・分野：関連すると思われる学問分野

具体例を図 3 に挙げる。なお、ワークショップでは、随時検索タスク、言語横断検索タスク、用語抽出・役割分析タスクの三つのタスクがあるが、本稿では、情報検索系のタスク、すなわち、随時検索タスクと言語横断検索タスクを対象として議論を進めている。

#### 4.2 NTCIR-1 におけるレlevance判定

レlevance判定にあたって、旧来のテストコレクションでは、トピックのみに注目したシステム志向のレlevanceを用いてきた。しかし、検索のシステムや手法は、変化してきており、それともなって、テストコレクション、およびテストコレクションにおけるレlevance判定も対応して変化していくべきであろう。すなわち、3 で述べたように、複合的で動的なレlevanceの特性をテストコレクションという枠組みのなかでどのように

図 3 検索課題の例

<p>&lt;検索課題 q=0069&gt;</p> <p>&lt;タイトル&gt; コンピュータ利用授業 &lt;/タイトル&gt;</p> <p>&lt;検索要求&gt; 小学校の教科におけるコンピュータの補助的利用について述べた文献。 &lt;/検索要求&gt;</p> <p>&lt;検索要求説明&gt; 情報化社会と言われるなか、コンピュータ利用能力を高めることが主張されており、特に幼年の時期からのコンピュータ教育の必要性が説かれる。小学校の教育実践でも、授業の中に補助教材のような形でコンピュータを活用することが行われている。そこで、小学校の教科の時間において、コンピュータが補助的に用いられている現状を知りたい。コンピュータを補助的に利用した教育実践の実際例が示されたもの。教科は問わないが、授業にどういった形でコンピュータが使われているか、具体的に説明されている必要がある。また、開発された教材ソフトを説明したものであっても、現場での利用の様子に言及されていれば正解。教科学習のためにコンピュータを補助的に用いるのでなければ不正解。コンピュータリテラシーを主眼に置き、教科教材との関連性がないものは不可とする。また、単に電子メールを使って他の学校の生徒とやりとりを行ったことを紹介したのも不可。 &lt;/検索要求説明&gt;</p> <p>&lt;概念&gt; a. CAI, computer assisted instruction, b. 学習指導, 授業, teaching, c. 教材, teaching material, d. 教育メディア, education media, e. マルチメディア, multimedia &lt;/概念&gt;</p> <p>&lt;分野&gt; 1. 電子・情報・制御, 8. 人文・社会 &lt;/分野&gt;</p>
---

反映させていくか、という問題である。しかし、一方では、これまでのテストコレクションとしての要件は満たさなくてはならない。

NTCIR-1 では、テストコレクションとしての要件を満たしつつ、俯瞰的な視点から捉えたレlevanceの多様性、複合性をできるだけ取り入れることを試みた。Mizzaro が述べている、レlevance判定に関係する次の五つの次元[10]にそって説明していく。

- (i) 判定するレレバンスの種類
- (ii) 判定者の種類：利用者、非利用者
- (iii) 判定に用いることのできる要素 I：文献表象要素、文献、情報
- (iv) 判定に用いることのできる要素 II：検索式、検索要求、情報ニーズ、問題状況
- (v) 判定する時点（時間）

#### 4.3 (i) 判定するレレバンスの種類

利用者志向のレレバンスに基づく判定では、I は I<sub>3</sub> に近く、U は U<sub>4</sub> に近く、D は D<sub>2</sub>、D<sub>3</sub> の割合が多いようなレレバンスを用いるが適当とされよう。しかし、テストコレクションには、個々の利用者の、システムとの対話のなかで推移していく性質を取り込むことは難しい。そこで、これまででは、システム志向のレレバンス、すなわち、I は I<sub>1</sub> に近く、U は U<sub>1</sub> に近く、D は D<sub>1</sub> の割合が多くなるようなレレバンスを用いてきた。

NTCIR-1 では、システム志向のレレバンスに、利用者志向のレレバンスの特徴を可能な範囲で取り入れるため、レレバンスをシステム志向と利用者志向とを両極とするスペクトラムで捉えた。NTCIR-1 では、このスペクトラムにおいて、「システム志向に近いところから利用者志向に歩み寄った地点」にレレバンスを位置づけている。

#### 4.4 (ii) 判定者の種類

テストコレクションの正解判定においては、利用者本人が判定することも可能であるが、そうすることが好ましいとは限らない。本人に依存する要素、すなわち、D<sub>2</sub>、D<sub>3</sub> や U<sub>3</sub>、U<sub>4</sub> などを必要以上に取り入れた判定では、システムを客観的数値で評価することが難しく、テストコレクションが満たすべき、妥当性や再現性の観点からは、適当とはいえない場合もある。例えば「既に読んだことがある文献はいらない」「著名な研究者の論文が欲しい」などという要素を判定に用いる場合は、判定は本人にしかできず、主観的な判定になる。

しかし、現実の情報検索におけるように、本人が判定者となった場合に意味を強く持つ変数も無視するわけにはいかない。そこで、「明文化できる

状況」つまり「他者が理解できる状況」を判定に取り込むことを考えた。具体的には、判定の助けとなり、一貫性を高めるために、コメント等を付したかたちで検索課題を収集し、これを用いて、第三者である他の研究者が複数人で複数回、判定を行なった。判定者に本人を含む場合は、判定者どうし話し合いを通して、検索要求と判定基準を可能な限り客体化するよう努めた。

NTCIR-1 では、検索課題中の検索要求説明の記述に、検索要求の背景や検索の目的など、すなわち、利用者の持つ情報要求(U<sub>3</sub>)や問題状況(U<sub>4</sub>)、および関心あるテーマ(D<sub>1</sub>)や取り組むべきタスク(D<sub>2</sub>)などを取り込んだ。これによって、第三者が理解して、あたかも本人の置かれた状況にいるかのようにふるまえる、いわば「擬似的状況 (pretended situation)」を構築した。

これらがないと判定が困難な場合もある。例えば、図3の課題では、検索要求説明から「実践例を探している」ことが判るので、小学校の教科におけるコンピュータの補助的利について述べていても、実践例を含んでいなければレレバントではない。また、「データマイニングの手法の最新動向が知りたい」という検索要求があったとしても、理論を研究しているのか、手許のデータを使って実用化を目指しているのかによって、レレバントな文書は異なる（同じ理論でも数学的な興味なのか、工学的な観点を重視するのか、などによっても違いが出てくる）。さらに、同じ研究者があるテーマで文献を探している場合でも、研究の初期段階でサーベイ的な論文を探すのと、研究が特定の手法や対象について実験や調査をした結果が知りたいのとでは、レレバントな文献は異なってくる。

#### 4.5 (iii) 判定に用いることのできる要素 I

基本的には I<sub>1</sub> であるが、NTCIR-1 の元データになっている「学会発表データベース」には、要約および著者キーワードが含まれており、これを擬似的に I<sub>2</sub> であると見なし、そこからくみ取れる I<sub>3</sub> も判定に用いた。これは、その分野にある程度知識のある者が判定を行うことで可能とした。

#### 4.6 (iv) 判定に用いることのできる要素 II

システム志向の変数だけでなく、利用者志向の変数も取り入れられるようにした。すなわち、基本的には  $U_2$  を用いるが、その背景にある  $U_3$  や  $U_4$  も、可能な範囲で判定に使用することを考えた ( $U_{1-4}$  は情報探索行動の連続した要素であり、もともと完全には切り離せないものである)。

そこで、NTCIR-1 では、検索課題を実際の研究者から収集し、その際、検索要求の背景や検索の目的などを記述したコメントを付してもらった。検索課題では、検索質問( $U_2$ )は「検索要求」として、コメントに含まれる情報要求( $U_3$ )や問題状況( $U_4$ )は「検索要求説明」として取り込んだ。

#### 4.7 (v) 判定する時点 (時間)

テストコレクションは、判定する時点は一点に固定される (逆に、一点に固定しないとレバンス判定が困難となる)。しかし、対話型システムの評価に利用することを考えている。すなわち、検索要求説明において利用者の模擬的な状況を示すことで、対話的システムへの利用の可能性を残した。なお、最近、対話型システムにおける、利用者志向 (主観) とシステム志向 (客観) のいずれか一方に偏らないかたちでレバンス判定を行う枠組みについての提案もなされつつある [11]。

検索システムの検索性能は、利用者とシステムの両方を含む空間で捉えるべきである。擬似的状況を判定者に提示することで、複数の同一状況の利用者の検索結果を比較することを提案している。

なお、(i)~(v)以外にレバンス判定に用いるスケールも判定結果に関係する要因となる。NTCIR-1 では、「レバント (A 判定)」「部分的レバント (B 判定)」「非レバント (C 判定)」という 3 値で行なっている。

## 5 おわりに

今後、本稿で述べたようなレバンスをめぐる枠組みの妥当性を実験的な側面も含め、検証していくことが必要である。さらに、多様な観点からレバンス判定を行なったテストコレクションの

開発も視野に入れていきたい。

NTCIR は、日本学術振興会未来開拓学術研究推進事業「高度分散情報資源活用のためのユービキタス情報システムに関する研究」のサブプロジェクト、「情報検索システム評価用テストコレクション構築」プロジェクトにより開発されている。本研究の成果は同プロジェクトによる。

## 注・引用文献

- [1] URL: <http://www.rd.nacsis.ac.jp/~ntcadm/>.
- [2] Kando, N. et al. "NTCIR: NACSIS Test Collection Project" [Poster] The 20th Annual Colloquium of BCS-IRSG, March 25-7, 1998, Autrans, France.
- [3] Kageura, K. et al. "NACSIS Corpus Project for IR and Terminological Research" Natural Language Processing Pacific Rim Symposium '97, Dec. 2-5, 1997, Phuket, Thailand, p. 493-6.
- [4] 神門典子 "情報検索システムの評価を巡って: テストコレクションとコンペティションを中心に" 1999 年情報学シンポジウム講演論文集, 1999 年 1 月 13-14 日, 東京, p. 129-36.
- [5] Voorhees, E. M. "Variations in Relevance Judgements and the Measurement of Retrieval Effectiveness", Proc. of the 21st Annual Int'l ACM SIGIR Conf. on R. and D. in Info. Retrieval, Aug. 24-28, 1998, p. 315-323.
- [6] 日本図書館学会用語辞典編集委員会編. 図書館情報学用語辞典. 丸善, 1997, p. 137.
- [7] ASK については、次の文献などを参照。  
- Belkin, N. J. et al. "ASK for Information Retrieval" J. of Documentation, 38, 1980, p. 77-112.
- [8] Barry, C. L. & Schamber, L. "User-defined Relevance Criteria: A Comparison of Two Studies" Proceedings of the 58th Annual Meeting of ASIS, Oct. 9-12, 1995, Chicago, Illinois, vol. 32, p. 103-11.  
利用者志向のレバンスは、次の文献などを参照。  
- Schamber, L. et al. "A Re-examination of Relevance: Toward a Dynamic, Situational Definition" Info. Processing & Management, 26, 1990, p. 321-43.  
- Barry, C. L. The Identification of User Relevance Criteria and Document Characteristics: Beyond the Topical Approach to Information Retrieval. Dissertation, Syracuse Univ., 1993.  
- Schamber, L. Users' Criteria for Evaluation in Multimedia Information Seeking and Use Situation. Dissertation, Syracuse Univ., 1991.
- [9] Saracevic, T. "Relevance Reconsidered '96" Proceedings Second Int'l Conf. of Library and Info. Science, Oct. 13-16, 1996, Copenhagen, Denmark, p. 201-8.
- [10] Mizzaro, S. "Relevance: The Whole History" JASIS, vol. 48, no. 9, Sep. 1997, p. 810-32.
- [11] 次の文献では、利用者の状況を記述し、それを用いて他者が検索する試みをしている。  
- Borlund, P. & Ingwersen, P. "Measures of Relative Relevance and Ranked Half-Life: Performance Indication for Interactive IR" Proceedings of the 21st Annual Int'l ACM SIGIR Conf. on R. and D. in Info. Retrieval, Aug. 24-28, 1998, Melbourne, Australia, p. 324-31.