社会技術研究論文集　Vol.3, 102-110, Nov. 2005

コミュニティオントロジーを利用した情報検索

INFORMATION RETRIEVAL USING COMMUNITY ONTOLOGIES

尾暮　拓也[1]・中田　圭一[2]・古田　一雄[3]

[1]科学技術振興機構 社会技術研究開発センター　(E-mail:ogure@diras.q.t.u-tokyo.ac.jp)
[2]東京大学大学院 新領域創成科学研究科 人工環境学専攻　(E-mail: nakata@k.u-tokyo.ac.jp)
[3] 東京大学大学院 工学系研究科 システム量子工学専攻　(E-mail: furuta@q.t.u-tokyo.ac.jp)

リスクを伴う様々な科学技術の賛否について社会的な合意を形成するためにはリスクコミュニケーションの促進が必要であるが，これを阻む大きな要因として専門用語や専門的概念の特殊性と難解さが挙げられる．本研究ではこの問題の解決策として情報検索の観点から専門家の概念体系を明示化して利用する手法を提案する．具体的には専門家の概念体系をコミュニティオントロジーとして定義し，検索質問に対する文書の適合度を計算するために用いる特徴ベクトルをオントロジーベクトルとして定式化する．そして実験を通じて提案する手法の検索性能の評価を行い，考察においてこの技術が市民に提供される場合の効果を論考する．

キーワード：リスクコミュニケーション，情報検索，専門家，コミュニティオントロジー，オントロジーベクトル

1. はじめに

現在の科学技術は専門領域の細分化が進んでおり，それぞれの分野の専門家がそれぞれに高度化した専門的知識を維持管理している．一方で「遺伝子組み換え作物のリスク」,「原子力のリスク」や「地球温暖化のリスク」など科学技術がもたらすリスクに対する社会的な関心も高まっているが，それぞれの専門分野で用いられる用語の特殊性と概念の難解さから，非専門家である一般市民がこれらの専門的知識に直接アクセスすることは容易ではない．**リスクコミュニケーション**を通じてリスクへの対応に係る社会的な**合意形成**を行うためにはこの敷居が最大の障害になると考えられる．

ここでもし身近にそれぞれの分野の専門家がいれば，非専門家はその人に相談することによって必要な専門的知識を得てリスクの理解を深めることができると考えられる．しかし身近に専門家がいない場合にはリスクの理解のために場当たり的に書誌を読むといったような能率が悪く危険なアプローチを取らざるを得なくなる可能性があり，これは深刻な問題である．したがって現在の社会では非専門家による専門的知識の理解を専門家の代理として支援する情報処理技術が必要とされているといえる．本研究ではこの問題に対して**情報検索**の方向からの解決方法を提案する．

一般的に**情報要求**(Information Need)を持つ人がそれを自己で正確に認識し，表現することは必ずしも容易でないとされる[1)]が，特に高度に技術的な事柄に関して非専門家が不安を持つ場合にはその情報要求は明瞭ではない可能性が極めて高い．このような漠然とした疑問を専門家に相談したときの専門家の思考プロセスを考えてみると，これは経験的にまず対話の中でその非専門家の問題意識を形成する概念を同定し，次にその概念を専門的な**概念体系**にマッピングして解釈するものと思われる．そして専門家はこの解釈に基づいて適切な専門的知識を回答したり，関連する書誌を推薦したりという対応を行っていると考えられる．従って専門家の代理としてリスクコミュニケーションを支援する情報検索技術を実現するためには概念体系，すなわち**オントロジー**を明示的に導入する必要があるといえる．

1.1. コミュニティオントロジー

専門家の概念体系はそれぞれの専門家の数だけ多様に存在すると考えられるが，専門家のグループを適切に切り出せばそこに共通的な概念体系が見出せる可能性がある．Lacher らは同一のコミュニティのメンバーの間ではコミュニケーションが円滑に行われている点に注目し，これはメンバーの間で実世界を観察する共通の観点を共有し，この共通の観点が共通の概念と語彙，及び概念の共通の体系化を規定しているためだと考察した[2)]．そしてこのようにコミュニティごとに共有されている概念体系を**コミュニティオントロジー**(Community Ontology)と呼んだ．社会技術研究の文脈で専門家の概念体系はここ

で議論されたコミュニティオントロジーと実質的に同一のものである。そこでこの論文ではコミュニティの共通認識として共有される概念体系をコミュニティオントロジーと呼ぶことにする.

1.2. 情報検索

図書の検索は古典的には図書館員によって**図書分類法**[4]などを利用して行われるものであり，**情報検索**[5]とはこの図書検索の計算機支援を指す．情報検索は近年では計算機性能の向上から全文の中に特定の文字列を含む文書を探す**全文検索**が実現され，「Google」など Web 上の検索サイトなどで広く利用されている．一方で Web の検索サービスでは他に図書分類と同様の旧来の分類階層に頼る**ディレクトリ型検索**と呼ばれるサービスがある．ところでこれらの分類階層は書誌の主題となる概念を整理した体系を与えるという点でオントロジーの一種であるとされ，専門家のコミュニティオントロジーをディレクトリ型検索の分類階層に応用して専門的知識を検索することも可能であるように思われる．一方で全文検索型の検索方法では明示的にオントロジーが考慮されておらず，適切な検索キーワードを検索者が自ら指定できなければならない．このような特徴は非専門家による専門的知識へのアクセスという条件には適切ではないと思われる.

ディレクトリ型検索サービスとしては「Yahoo!」や「Open Directory Project」などがあるが，これらのサイトでは図書館での図書分類法と同様に人間が Web サイトの主題を主観的に判断してハイパーリンク集を作成している．この方法の長所は，利用者が自ら疑問を持つ概念の位置づけや呼び方を知らなくても人手によって作成されたオントロジーを案内にして探索できる点であり，短所はサービス提供側がリンクを一つ一つ編集しなければならないので人的コストが高い点である．特に科学技術の各分野ではコミュニティが独立に発達しているので，それぞれのコミュニティオントロジーに対応するディレクトリ型検索サービスを構築すると非常に高いコストがかかる．また特定技術に対する肯定派，反対派など同一分野に複数のコミュニティのオントロジーが存在する場合には同一分野にさらに複数のディレクトリ型検索サービスを構築する必要ができてしまう．このコストの問題を解決するためには，コミュニティオントロジーが与えられたときに検索対象となるホームページのインデックスを自動で作成して検索可能にする技術が必要であると思われる．そこで本研究では専門家のコミュニティオントロジーを利用した自動的な情報検索手法を提案する.

1.3. 検索性能について

情報検索の分野では与えられる**検索質問**を自動的，あるいは半自動的に改良する**検索質問拡張**(Query Expansion)の研究が行われており， Baeza-Yates らはここで用いられる手法を次のように分類した[6].

1. **適合性フィードバック**
   検索結果の文書をユーザーが閲覧して適合性の判断を行い，これを次の検索にフィードバックする手法.
2. **自動ローカル分析**
   検索時に内部的に仮検索を実施して結果を分析し，これを基に検索質問を改良して本検索を行う手法.
3. **自動グローバル分析**
   あらかじめ検索対象文書を分析してユーザーの検索質問を解釈するための情報を用意しておく手法.

ここで自動グローバル分析によって得られる情報は通常，語彙の間や文書の間の関連性であり，我々が取り扱うオントロジーと同様の情報であると考えられる．しかし Baeza-Yates らはこの自動グローバル分析は単純な同義語補完のための利用を除いてはパフォーマンスの向上に寄与しないものと考えられてきたとしている．この理由として実際にユーザーが期待する検索質問の解釈のされ方は検索が必要となった状況への文脈依存性が高く，事前に文書を分析して得られる情報ではユーザーが検索質問を発行するに至った文脈まではカバーできないからであると考察している．また Salton[7][8]は実験を行い，同義語の情報を与える**シソーラス**を用いた語彙論的な手法は検索性能を向上させるものの，オントロジーのように**上位概念**，**下位概念**といった知識構造に踏み込んだシソーラスはパフォーマンスをそれほど向上させないと結論付けた. その原因として Baeza-Yates らと同様にユーザーの目的や要求の多様性を指摘した.

本研究では専門家のコミュニティオントロジーを利用した情報検索手法を提案しているが，この枠組みでは適切なオントロジーをユーザーに選択させることによってユーザーが情報検索を望む文脈を検索質問へ反映できることが期待できる．一方で当該コミュニティの専門家による評価でパフォーマンスに改善が見られない場合は，この手法の主眼である「検索質問の専門家の視点からの解釈」に失敗している可能性を示す．そこで本研究では専門家コミュニティのオントロジーを利用してその専門家コミュニティのメンバーが検索を行う場合を想定した実験を行い，専門的な情報検索におけるコミュニティオントロジーの有効性を検証する．そして考察において専門的知識を持たない市民が専門家のコミュニティオントロジーを利用する場合の有効性について議論する.

2. オントロジーとテストコレクション

特定の技術的なコミュニティのメンバーにとってのオ

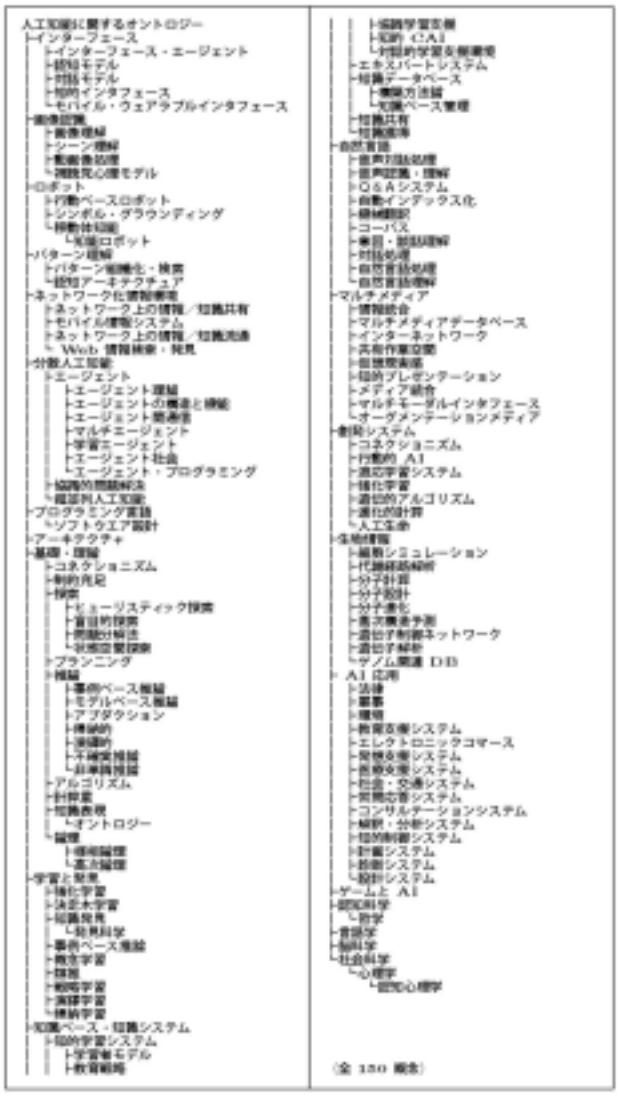

Fig.1 人工知能研究分野のトップレベルオントロジー

ントロジーを用いた情報検索の性能を評価するためには，そのコミュニティのオントロジーを明示的に記述したデータが必要である．また評価用の検索課題，いわゆる**テストコレクション**も用意しなければならない．ここでこの実験で対象とするコミュニティには人工知能分野のコミュニティの一つとして筆者が所属する大学の研究室を選んだ．コミュニティオントロジーは以下のように構築された．

まず，日本人工知能学会から提供される投稿論文の「該当分野」の表 [9]を参考にしてメンバーで討論し，最大 4 階層の階層構造で 150 概念からなる人工知能の**トップレベルオントロジー**(Fig.1)に合意した．次に，人工知能の書籍 [10]の日本語の巻末索引 2,209 語を参照し，このトップレベルオントロジーの適切な概念の下位に配置した．配置は書籍の文章の中で索引語が出現する文脈を参考に行われ，「情報処理」などトップレベルオントロジーの中に適切な上位概念がないと判断された索引語は無視された．結果として索引語 1,063 語が新たな概念ノード及びその概念を指示する専門用語の語彙の情報として追加された．また後述の検索質問に含まれる内容語のうちここでカバーしきれなかった 13 語について，これらが存在しないと検索実験が不可能であることから例外的に追加された．複数の異なる文脈で同じキーワードが用いられる場合があり，この場合は異なる概念を指していると解釈して重複してオントロジーに登録した．結果として 1,431 概念からなる人工知能分野のコミュニティオントロジーが構築された．ただし同義語と思われるキーワードを同一の概念にまとめる処理は行われていない．この作業のために専用のソフトウエアツールが開発された [11]．

テストコレクションとしては，日本語のものでは NTCIR プロジェクト [12]によって NTCIR-1 が構築され公開されている．NTCIR-1 には 33 万の論文のアブストラクトからなる検索対象文書と 83 の**検索課題**，及び検索課題に対する正解情報が収録されているが，この中に日本人工知能学会から論文アブストラクトが 2,031 文書提供されている．またこの 83 の検索課題には「ソーラーカー」など人工知能分野とは直接関係がないと思われる検索質問で表現されたものも含まれる．人工知能分野の専門的な情報検索手法評価用のテストコレクションとして，NTCIR-1 の検索対象文書の中で人工知能学会から提供された 2,031 文書サブセットと，83 検索課題のうちの人工知能分野の検索性能評価のために適切と思われる検索課題サブセット 10 課題を選択し(Table 2)，さらにこれらに対する検索結果として有効かどうかの正解情報を 2,031 文書全てに対してコミュニティのメンバーの手作業により評価して用意した．以降ではこの NTCIR-1 のサブセットに新たに正解判定を行ったテストコレクションを FTCIR(Furuta lab. Test Collection for IR systems)と呼ぶ．この FTCIR の概要を Table 1 に示す．また FTCIR と既成のテストコレクションとの比較を Table 3 に示す．この比較から他のテストコレクションと比較して FTCIR は検索質問の数がやや少ないことが分かる．これは FTCIR を他

Table 1　テストコレクション「FTCIR」の概要

適用条件:	検索者が人工知能コミュニティオントロジーを知っている場合の人工知能分野の情報検索.
文書集合:	2,031 文書 (NTCIR-1 中の日本人工知能学会論文アブストラクト全て)
検索質問:	10 質問 (NT-CIR-1 より抜粋, 表 2)
正解判定:	2 名が分担, 1 つの文書に対し 1 人が判断.
判定基準:	A 判定:検索要求に完全に適合する. B 判定:検索要求に部分的に適合する. (NTCIR-1 に準拠)
正解数:	最大 123 文書, 最小 4 文書, 平均 35.0 文書 (A 判定+B 判定)

Table 2　　FTCIR に採用された NTCIR-1 の検索課題

質問番号	検索質問
0001	「自律移動ロボットについて」
0003	「機械学習におけるサンプル複雑性について論じている文献」
0006	「エージェント機能を利用した知的情報検索」
0007	「大規模なデータベースとユーザーとのインタラクションにおけるユーザーの認知的側面に関して論じている論文」
0012	「データマイニング手法を改良，提案している文献」
0014	「故障診断システムについて」
0028	「ニューラルネットワークの手法，理論，原理などについて記述した文献がほしい.」
0056	「設計・製造・保全などの人工物のライフサイクルの異なる部門間での情報共有および知識共有について報告した論文・記事が欲しい.」
0057	「設計やアイデア生成などの創造的思考について，その過程のモデル化および支援に関して述べた論文・記事が欲しい.」
0059	「シソーラスの自動的な構築方法，あるいは維持方法について述べた文献が欲しい.」

のテストコレクションのように単独で定量評価に利用することには危険があることを示すが，しかし今回の実験のように他のテストコレクションとの定性的な比較のために用いられるのであれば十分に有用な知見が得られると考えられる．

なおここで人間の恣意が影響する可能性を含む作業について，(1)「人工知能書籍の索引語のオントロジーへの追加」及び「検索課題の選択」と(2)「テストコレクション正解判定」の作業は別のメンバーによって行われ，これらのメンバー間にはコミュニティオントロジー以外に情報検索結果にかかわる情報は共有されていないので実験結果を操作する余地はない．また「索引語の追加」に関しては出典を限定したためにそもそも恣意の余地は小さいと思われる．

3.　検索アルゴリズム

本研究では**ベクトル検索モデル**で用いられる文書の**特徴ベクトル**をコミュニティオントロジー中の各概念の強度で表現した**オントロジーベクトル**を提案する．計算論的にはこのようなベクトル操作による検索質問拡張は既に多く提案されているが，この提案の本質は専門家コミュニティのメンバーによる文書の専門的な解釈をオントロジーベクトルとしてモデル化することにある．このモデル化によって異なるコミュニティのオントロジーに依れば同じ文書も異なる解釈が与えられるという現実の多様性を文書の機械的インデックスに反映することができ，これは情報検索技術にとって画期的な進歩といえる．この手法の枠組みを以下のように定式化する．

Table 3　既存のテストコレクションとの比較

コレクション	文書数	検索質問数
ADI	82	35
CACM	3,204	64
CISI	1,460	112
CRAN	1,398	225
MED	1,033	30
NPL	11,429	100
TIME	425	83
LISA	6,004	35
INSPEC	12,684	84
OHSUMED	348,566	106
CF	1,239	100
TREC-6	1,754,896	350
NTCIR-1 (original)	332,918	83
FTCIR	2,031	10

ステップ１：コミュニティオントロジーを明示化する

専門的な知識を共有するコミュニティのオントロジーを以下の形式で用意する．

オントロジーはネットワークグラフの形に体系化された概念ノードからなるとし，ここでは便宜的に概念と専門用語を区別して表現する．

$$S_G = \{Concept_1, Concept_2, \cdots, Concept_h\} \quad (1)$$

S_G はコミュニティ G で共有されるコミュニティオントロジーの要素である概念の集合であり，$Concept_i$ はその概念である．

$$\mathbf{C}=\begin{bmatrix} 0 & c_{12} & \cdots & c_{1h} \\ c_{21} & 0 & \cdots & c_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ c_{h1} & c_{h2} & \cdots & 0 \end{bmatrix} \quad (2)$$

$\mathbf{C}$ は概念のネットワークグラフを隣接行列で表現したものであり，c_{ij} は$Concept_i$から$Concept_j$へリンクがあるときにそのリンクの属性に応じた**活性伝播率**である．活性伝播率については後で議論する．

$$T_G=\{t_1,t_2,\cdots,t_n\} \quad (3)$$

T_G はコミュニティ G で用いられる専門用語の集合であり，t_i はその専門用語である．

$$\mathbf{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1h} \\ p_{21} & p_{22} & \cdots & p_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nh} \end{bmatrix} \quad (4)$$

P は専門用語と概念との対応を表現した行列である．

$$p_{ij}=\begin{cases} 1(t_i\text{が}concept_j\text{を意味するとき}) \\ 0(\text{それ以外}) \end{cases} \quad (5)$$

ステップ2：文書の単語ベクトルを計算する

検索対象となる文書 k の中に専門用語 t_i が出現する頻度 d_{ki} を数えて単語ベクトル $\mathbf{d}_k$ を計算する．このベクトルは原始的なベクトル検索モデルで用いられる特徴ベクトルである．

$$\mathbf{d}_k={}^t\left(d_{k1},d_{k2},\cdots,d_{kn}\right) \quad (6)$$

ステップ3：単語空間からオントロジー空間に写像する

単語空間から概念空間への変換を行う行列 $\mathbf{M}$ により文書のオントロジーベクトル $\mathbf{D}_k$ を求める．ここでは**活性伝播モデル**(Spreading Activation Model)[13) 14)]に準じた単純なモデルを用いる．活性伝播モデルとは人間の連想記憶をモデル化したものであり，**意味ネットワーク**上で刺激を受けた概念ノードが起点となって近接するノードに**活性化**(activation)と呼ばれる想起の程度を記入していくものである．ノードの間のリンクの**結合強度**(accessibility)を本研究では**活性伝播率**と呼ぶことにし，この活性伝播率 c_{ij} を[0，1]の範囲で定義する．ある単語ベクトル $\mathbf{d}_k$ が与える起点から活性伝播を行ってｉステップ目に活性伝播する概念の活性化 $\mathbf{a}(i)$ は，

$$\mathbf{a}(i)={}^t\mathbf{C}^i\,{}^t\mathbf{P}\,\mathbf{d}_k \quad (7)$$

であることから，λ ステップ目まで活性伝播を行う場合には，文書のオントロジーベクトル $\mathbf{D}_k$ はこれらを重ね合わせて

$$\mathbf{M}=\left(\mathbf{I}+{}^t\mathbf{C}+{}^t\mathbf{C}^2+\cdots+{}^t\mathbf{C}^{\lambda}\right){}^t\mathbf{P} \quad (8)$$

なる変換行列Ｍについて，

$$\mathbf{D}_k=\mathbf{M}\,\mathbf{d}_k \quad (9)$$

となる．ここで $\mathbf{I}$ は単位行列である．オントロジーのグラフに循環がある場合は連想のステップ λ について合理的な値を検討しなければならないが，今回の実験で用意したオントロジーは木構造であり、λ をオントロジー階層の最大の深さとすれば収束する。

ステップ4：適合度を計算する

検索質問がオントロジーベクトル $\mathbf{Q}$ で与えられた場合，検索質問への文書 k の適合度 $\mathrm{Sim}(\mathbf{Q},k)$ は一般的なベクトル検索モデルと同様に計算される．

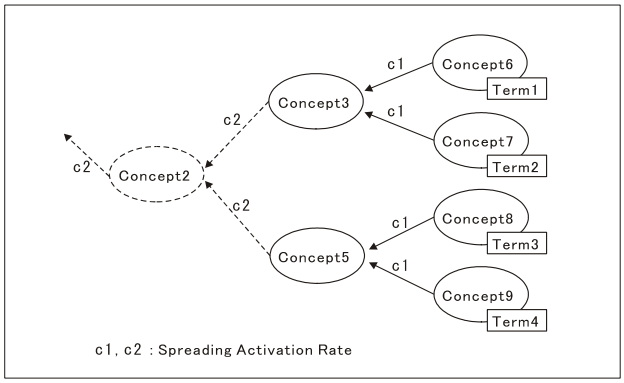

Fig.2 活性伝播率 (c1, c2) の設定

$$\mathrm{Sim}(\mathbf{Q}, k) = \cos\theta = \frac{\mathbf{Q} \cdot \mathbf{D}_k}{\|\mathbf{Q}\| \|\mathbf{D}_k\|} \qquad (10)$$

今回の実験では検索質問が文章で与えられるので$\mathbf{Q}$は文書のオントロジーベクトルと同様に求められる.

4. 実験方法

「FTCIR」の「B 判定基準」を用いて以下の条件で実験を行い，オントロジーを利用した場合の検索性能を比較する.

(a) 単語の特徴ベクトルによる検索.
(b) 弱構造化シソーラスを利用した特徴ベクトルによる検索.
(c) コミュニティオントロジーによるオントロジーベクトルを特徴ベクトルにした検索.

性能の定量には一般的に使用される指標である **11 点平均精度**を用いる．(a)の条件は検索質問拡張を全く行わない原始的な検索性能を与える．(b) の条件は同義語シソーラスを用いる手法などオントロジーの高次構造を考慮しない検索質問拡張による検索性能を想定したものであり，一般的な方法で改良された性能を与える．ここでいう弱構造化シソーラスとは同義語，類義語や比較的限定されたトピックで出現する共通の語彙をまとめたものとする．(c)の条件は本研究で提案するオントロジーによる支援により改良された性能を与える．活性伝播率は問題の簡略化のため Fig. 2 のように 2 つの平均値に縮退できると仮定する。両者を区別する根拠はオントロジーの構築時にそれぞれのリンクが発見された経緯の違いである。ただしここで平均値を用いる点については問題の簡略化以外の正当化はなく、実際に概念間の距離の違いを推定するさまざまな研究があるので改良の余地がある。

比較する条件について具体的には以下のとおりである。(a)では用意した人工知能オントロジーが保持している専門用語の語彙 1,076 語の単語ベクトルを用いる．(b)では用意したコミュニティオントロジーのトップレベルの構造を無視したオントロジーベクトルを用いる．これは Fig. 2 に示すオントロジーの上位の概念への活性伝播率 c2 をゼロとすることに等しい．チューニングパラメータは c2 のみであり{0.1, 0.2, …, 1.0}の範囲でサーベイする．(c)では原始的な活性伝播モデル [13]に習って上位概念に

Table 5　11 点平均精度による性能評価

検索条件	11 点平均精度
FTCIR	
(a) 単語	0.187
(b) 弱構造化シソーラス(c1=0.7)	0.247
(c) オントロジー(c1=0.7, c2=0.3)	0.288
NTCIR-1A	
(a) 単語	0.213
(b) 弱構造化シソーラス(c1=0.7)	0.300
(c) オントロジー(c1=0.7, c2=0.1)	0.309

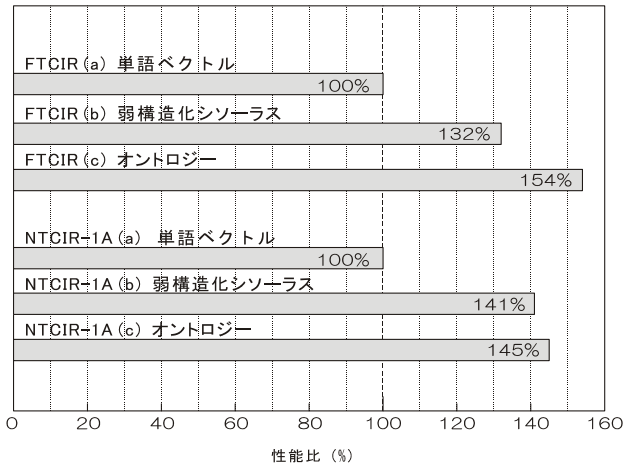

Fig. 3　単語による検索 (a) を基準とした性能比

Table 4　NTCIR-1 の対照実験用サブセット「NTCIR-1A」

適用条件:	一般
文書集合:	2,031 文書 (NTCIR-1 中の日本人工知能学会論文アブストラクト全て)
検索質問:	10 質問 (FTCIR と同一)
判定基準:	A 判定:検索要求に完全に適合する. B 判定:検索要求に部分的に適合する.
正解数:	最大 30 文書, 最小 1 文書, 平均 11.5 文書 (A 判定+B 判定)

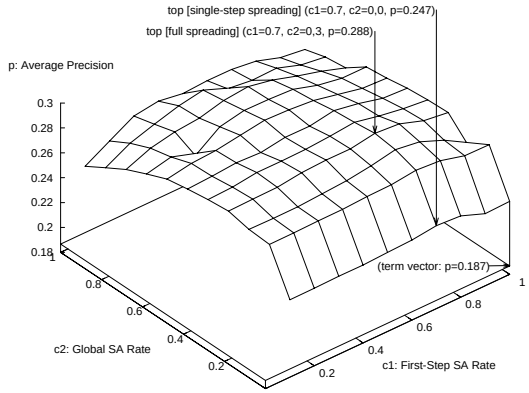

Fig.4　FTCIR の活性伝播率のパラメータサーベイ

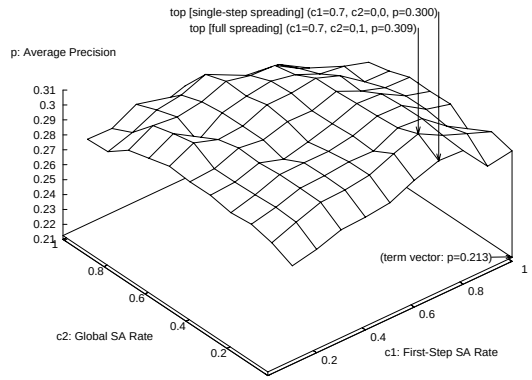

Fig.5　NTCIR-1A の活性伝播率のパラメータサーベイ

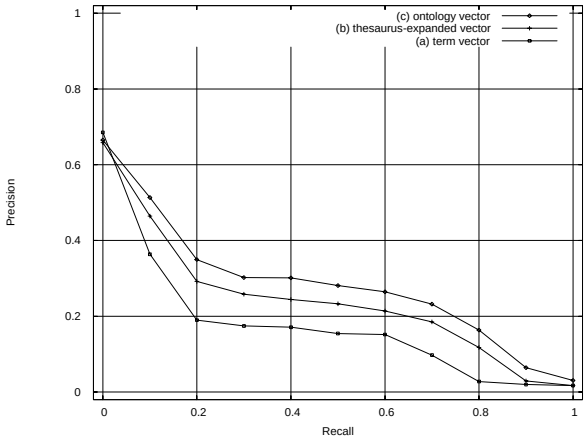

Fig.6　FTCIR の 11 点平均精度グラフ

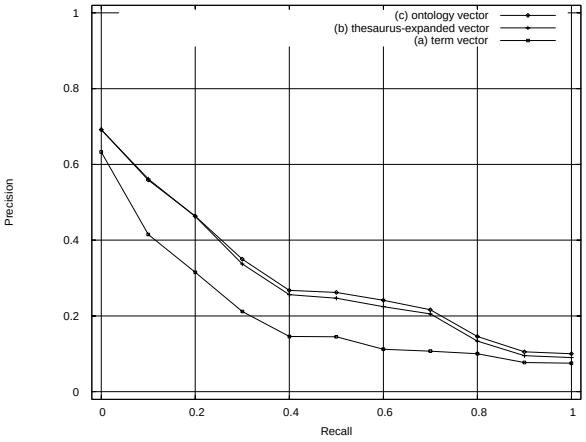

Fig.7　NTCIR-1A の 11 点平均精度グラフ

伝播する活性化のみを考え，Fig. 2 の活性伝播率 c1 と c2 をチューニングパラメータとしてそれぞれ{0.1, 0.2, …, 1.0}の範囲でサーベイして最大検索性能について比較する．対照実験として FTCIR と条件を揃えた NTCIR-1 のサブセット NTCIR-1A を用意し(Table 4)，人工知能分野のコミュニティオントロジーを前提としない正解情報に対する検索性能と比較する．

5.　実験結果

標準的な検索性能指標である 11 点平均精度の比較を Table 5 に示す．活性伝播率のパラメータサーベイについて FTCIR のサーベイの結果を Fig. 4 に示す．なお図中において縦軸は 11 点平均精度である．(b)の条件では c1 = 0.7 の時に最大性能となった．また(c)の条件では c1 = 0.7, c2 = 0.3 の時に最大性能となった．NTCIR-1A についてのパラメータサーベイの結果を Fig. 5 に示す．(b)の条件では c1 = 0.7 の時に最大性能となった．また (c)の条件では c1 = 0.7, c2 = 0.1 の時に最大性能となった．それぞれの条件での 11 **点平均精度グラフ**を Fig. 6，Fig. 7 に示す．さらに単語による検索(a)の性能を基準として弱構造化シソーラスを用いた場合(b)の性能とオントロジーを用いた場合(c)の性能を正規化した性能比を Fig. 3 に示す．

6.　考察

6.1.　実験結果について

Fig. 3 より，NTCIR-1A では単語のベクトルモデル(a)に対して弱構造化シソーラスを適用した検索(b)は性能が向上した．オントロジーベクトルによる検索(c)では弱構造化シソーラスによる検索から大きな性能向上は見られなかった．しかし FTCIR ではオントロジーベクトルによる検索で大きく性能が向上しており，FTCIR の条件ではコミュニティオントロジーの上位構造が検索性能を改善することが分かった．この傾向は Fig. 6，Fig. 7 の 11 点平均精度グラフを見比べても明らかである．従って NTCIR-1A と FTCIR とに対しては人工知能分野のコミュニティオントロジーの効果が明らかに違うといえる．

NTCIR-1 の正解判定を行った人物がどのようなコミュ

ニティに所属していたかは不明だが，NTCIR-1 全体に含まれる文書の分野は科学一般であるので，人工知能分野のみを意識して正解判定を行ったわけではないと考えられる．そこでこの人物が科学一般について興味を持つコミュニティを代表していると仮定すると以下の事が言える．すなわち NTCIR-1A の実験結果は「類義語シソーラスなどのローカルな情報に比べて領域全体の概念を体系化した情報は一般的な情報検索の性能向上に大きく寄与しない」という Baeza-Yates らや Salton の指摘(1.3 節)を支持するものと思われる．これに対し FTCIR の結果は，専門的なコミュニティオントロジーを利用する検索手法について，そのコミュニティに属する専門家の正解情報による評価を用いる場合に限って検索性能に程度の大きい改善が認められることを示す．これらから専門的文書に対する専門家のレレバンス判定を情報検索システムで再現する目的に限って専門家が所属するコミュニティのオントロジーを利用することが効果的であると結論付けられる．

6.2. 非専門家による情報検索について

非専門家が専門的な文書を検索する条件では，検索された文書のその人にとっての有効性すなわち**レレバンス**[15]の客観的評価が難しく，再現性のある評価実験は行いにくいが，以下に専門家の持つコミュニティオントロジーを利用した非専門家による情報検索についての論考を加える．例えば非専門家が自分の持つ疑問について専門家に相談したいと考えるような場合，その人は現在自分が持っている知識が不十分であると感じていると言われる．この状態は**情報要求**[1]と呼ばれ，頭の中にある情報要求は必ずしも明確に言語化できるとは限らないとされる．相談を受けた専門家は相談者との対話を通じてこの情報要求を推定し，情報要求を構成する概念を専門的に解釈するであろう．またこのとき相談者は情報要求が専門家の概念体系で解釈されることを期待しているものと思われる．例えば地震の危険性について専門家に相談するときには，地質学者に相談する場合と建築士に相談する場合とでは異なる解釈が行われることが予想できるし，いずれのコミュニティオントロジーによる解釈を希望するかによって情報要求の最初の明示化を行っていると考えられる．今回の実験条件では検索質問を文章で与えたが，提案する手法では検索者がオントロジーから情報要求に関連すると感じる概念を選択して検索システムに与える使い方が可能であり，この場合はオントロジーは情報要求の具体化のためのツールとなる．従ってもしオントロジーを利用した情報検索システムが専門家のレレバンス判定を再現できる場合，未知のオントロジーの中でのオリエンテーションの問題は依然として残るものの，非専門家がいずれかの専門家を選択して相談する代わりにいずれかのオントロジーを選択して情報検索を実行できることになり，社会的に有用であると考える．

この手法の問題点はコミュニティオントロジーを構築するコストの高さと，非専門家が見慣れない専門家のオントロジーを使って情報要求を表現しなければならない点であろう．前者についてはかねてより知識ベース構築のコスト問題が指摘されてきたが，近年は複雑化する科学技術の諸問題への理解を社会的に醸成するための知の体系化の必要性が社会技術の文脈で提唱されており，社会的に関心の高い専門分野のコミュニティオントロジーを構築することへの要請が高まりを見せている点ではこれまでの状況とは異なっている．後者については，専門家に相談する場合には専門家のオントロジーの構造まで考慮しなくてよい点が一つの利点であるので，本研究で提案する情報検索手法は専門家の完全な代替にはなれないことが指摘できる．ただしオントロジーが階層構造を持っていて，かつ各概念について十分な説明文が用意されている場合は，非専門家によるオントロジーの探索は専門家との対話と比較して必ずしも非効率なものとは思えない．またここで検索者が何らかのオントロジーを持っていることが予想される場合はオントロジーマッピングの研究[2]を応用するなどの手法が有効であると考えられ，今後はこのような非専門家の疑問の明示化を支援するための研究が期待される．

7. 結論

社会的関心の高い科学技術に係わる専門的な知識の非専門家による探索を支援するために，専門家のコミュニティオントロジーを明示的に内蔵した情報検索手法を提案した．この手法による検索結果が専門家のレレバンス判定に近いことを実験で示し，さらに専門家のレレバンス判定に近い検索結果を提供するシステムが非専門家に提供される場合の利点を論考した．

参考文献

1) 徳永健伸(1999)『言語と計算 5 情報検索と言語処理』 東京大学出版会.

2) Lacher, M. S., Wörndl, W., Koch, M., and Brede, H. (2000) *Ontology mapping in community support systems*. Dept. of Computer Science, Technische Universität München.

3) Goldman, A (2001). Social Epistemology. In Zalta, E. N. (Ed) *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2001/entries/epistemology-social/

4) 緑川信之(1996). 『本を分類する』勁草書房.
5) Frakes, W. B. and Baeza-Yates, R. eds (1992). *Information Retrieval: data structures and algorithms*. Prentice-Hall, Upper Saddle River, New Jersey.
6) Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. ACM Press, New York.
7) Salton, G. (1988). Automatic Indexing and Abstracting. In Willett, P. (Ed), *Document Retrieval Systems* (pp. 42–80). Laylor Graham, London.
8) Salton, G. (1998). 「自動テキスト分析」 In 上田修一 (Ed), 『情報学基本論文集 II』 (第 2 章, pp. 17‐47) 勁草書房.
9) 日本人工知能学会. 『人工知能学会誌原稿執筆案内』 http://www.ai-gakkai.or.jp/jsai/journal/toauthor.pdf
10) Russell, S. and Norvig, P. (2000). 『 エージェントアプローチ人工知能』共立出版.
11) Ogure, T., Nakata, K., and Furuta, K. (2001). *Ontology Processing for Technical Information Retrieval*. in Proc. of the 9th Int. Conference on HCI, Vol. 1 (pp. 1503–1507).
12) 神門典子(2002)「NTCIR とその背景」, 『 人工知能学会誌』, Vol. 17, No. 3 (pp. 296‐300).
13) Collins, A. M. and Quillian, M. R. (1969). *Retrieval Time from Semantic Memory*. J. Verbal Learning and Verbal Behavior, Vol. 8 (pp. 240–247).
14) Collins, A. M. and Loftus, E. F. (1975). *A Spreading-Activation Theory of Semantic Processing*. Psychological Review, Vol. 82, No. 6 (pp. 407–428).
15) Saracevic, T. (1998). 「レレバンス: 情報学におけるその概念の概観と枠組み」 In 上田修一 (Ed) 『情報学基本論文集 II』 (第 6 章, pp. 117‐183) 勁草書房.

謝辞

本研究は，社会技術研究開発センター　ミッション・プログラム「安全性に係わる社会問題解決のための知識体系の構築」(平成 13～14 年度は日本原子力研究所の事業，平成 15 年度からは独立行政法人科学技術新興機構の事業）の研究として行われた.

INFORMATION RETRIEVAL USING COMMUNITY ONTOLOGIES

Takuya OGURE [1], Keiichi NAKATA[2], and Kazuo FURUTA[3]

[1]Research Institute of Science and Technology for Society, JST (E-mail: ogure@diras.q.t.u-tokyo.ac.jp)
[2] Institute of Environmental Studies, The Univ. of Tokyo (E-mail: nakata@k.u-tokyo.ac.jp)
[3] QUEST, The Univ. of Tokyo (E-mail: furuta@q.t.u-tokyo.ac.jp)

It is unquestionable that “risk communication” is required for consensus building on many risks of technologies. The communication, however, can be obstructed by unfamiliarity and difficulty of technical terms or technical concepts which experts have developed for a long time. To improve accessibility to the technical information, we propose a novel technique for information retrieval (IR) which uses *community ontology* of experts. We will detail the notion of community ontology for formulating *ontology vector*, an ontology-oriented feature vector to introduce it into the conventional vector IR model. Then we will show results of an experiment demonstrating improvement of performance brought by the proposed technique, and discuss the effect that will be produced by the technique in practical use.

Key Words: *Risk Communication, Information Retrieval, Experts, Community Ontology, Ontology Vector*