社会技術研究論文集 Vol.5, 206-215, Mar. 2008

オントロジーを利用した分野特化型情報検索技術の社会的実装

SOCIAL IMPLEMENTATION OF ONTOLOGY-BASED DOMAIN-ORIENTED INFORMATION RETRIEVAL

尾暮　拓也[1]・古田　一雄[2]

[1]博士（工学）　産業技術総合研究所　知能システム研究部門　研究員　(E-mail: ogure.takuya@aist.go.jp)
[2]博士（工学）　東京大学大学院　工学系研究科　教授　(E-mail: furuta@q.t.u-tokyo.ac.jp)

専門的知識は非専門家の問題解決や意思決定のためにも利用できるように共有されることが望ましい．しかし原子力と社会の問題に目を向けると，その専門的知識の共有不足が社会的に議論を深めるための障害になっていると考えられる．これは知識の共有の不全がもたらす問題の顕在化であるといえる．そこで本研究では原子力分野を対象として，専門的なWebコンテンツを容易に検索できる分野特化型の情報検索技術を開発し，社会に実装した．この情報検索技術は対象とする専門分野に関するオントロジーを用意して活用するものである．この技術は専門的知識の流通支援のために幅広い応用が可能であると考えられる．

キーワード：情報検索，オントロジー，原子力安全，NOOCLE，専門的知識へのゲートウェイ

1. 序論

様々な専門分野で蓄積されている専門的知識は，一般市民の意思決定や問題解決のためにも有用であると考えられる．特に原子力発電など，環境への影響が大きい科学技術に関する意思決定は社会的な関心事であるといえる．しかし，一般市民がこのような科学技術に関する知識を得ようとすると「専門性の壁」が障害となる．この壁をうまく通り抜けるための方法を提供することは専門家の側の責任であると考えられる．そこで本研究では原子力分野を対象にして，「専門的知識へのゲートウェイ」を提供するための社会技術の開発に取り組む．

地球温暖化の進展などの社会情勢の変化により，近年ではエネルギー政策のひとつである原子力政策について関心が高まっている．原子力政策の可否については市民による理性的な議論が不可欠であるが，焦点になる「原子力安全」に関する技術体系は高度に専門的であるため，非専門家がその技術の有効性の質と程度を判断することが難しいという問題がある．他の様々な専門分野でも，このような「専門性の壁」の問題は同様に存在していると思われる．ここではまず，非専門家にとっての「専門性の壁」とは何かを分析してみることから始める．

一般に，人が何らかの課題に対処するために情報を必要としている状態を「情報要求」という．徳永は情報要求の定義を情報工学に立脚して「ユーザがある目的を達成するために現在持っている知識では不十分であると感じている状態」としている[1)]．さらにTaylorは図書館学に立脚して情報要求を以下のように段階わけしている[2)]．

- Q1 **直感的要求**：現状に満足していないことは認識しているが，それを具体的に言語化してうまく説明できない状態．
- Q2 **意識された要求**：頭の中では問題を意識できるが，あいまいな表現やまとまりのない表現でしか言語化できない状態．
- Q3 **形式化された要求**：問題を具体的な言語表現で言語化することができる状態．
- Q4 **調整済みの要求**：問題を解決するために必要な情報の情報源を同定できるくらい問題が具体化された状態．

Q1の「直感的要求」とは例えば，原子力問題に関していえば，原子力発電所に対して漠然と恐怖感を抱いているような状態であると考えられる．Q2の「意識化された要求」とは例えば，原発は危険なのではないか，というような問題意識を持った状態であると考えられる．Q3の「形式化された要求」とは例えば，「原子炉の放射能で環境が汚染されるリスクはどの程度か」というような言語化された問いが立てられる状態であると考えられる．Q4の「調整済みの要求」とは例えば，「生体組織の放射線感受性の資料はどこにあるか」や「レベル３-PSAの解析結果はどこにあるか」といった直接的な質問をすることが可能な状態であると考えられる．

Q1やQ2の段階ではまだ認識が暗黙的であり，内省や

対話を進めるしかないと考えられるが, Q3 の段階まで進むことができれば，問題に切り込むための「問い」が立つと考えられる．さらに Q4 の「調整済みの要求」の段階に達することができれば，既存の情報検索技術に頼って情報要求を解決することが可能になる．「専門性の壁」で問題となっているのは，Q2 の「意識化された要求」の段階から前に進めなくなってしまう場合であると考えられる．必要としている知識が専門分野の中で体系化されている場合，その専門分野の「ことば」を知らないと，「情報要求」そのものを自身でうまく認識できないままで思考停止してしまうと考えられる．これは「問い」が立てられない状態であるといえる．

専門分野の「ことば」を理解することが難しい理由の一つは，それらの指す内容の多くが「構成概念」[3]であるからであると考えられる．科学哲学において「構成概念」とは，ものごとを効率的に説明するために仮置きされた架空の概念であると説明される．「エネルギー」,「リスク」など，科学的な認識の枠組みをかたちづくっている概念の多くは，対応する実体がない「構成概念」である．構成概念は様々な分野で様々に設定されているが，これらはあくまでも架空であって，他の概念との関係性のみによって定位している．従って，これらのひとつひとつはその分野の「概念体系」の構造を知らなければ理解できないと考えられる．このように，一般市民は構成概念を使えないことから専門分野で扱われる事柄についての「認識の枠組み」を得にくく，有効な「問い」を立てられないと考えられる．この敷居が「専門性の壁」の正体であると考えられる．

「原子力安全」分野の専門家は，この専門性の壁に対処することにも責任を負っている．原子力安全の専門家に求められる社会参加の様式は，専門家のリスクの解釈を一方的に伝える「リスクコミュニケーション」から，リスクの議論を形成するための知識を提供する「リスクデリバレーション」に変化しつつある [4]．従って原子力の専門家には，リスクに係る意思決定に必要な専門的知識を一般市民が利用できるように提供する責任がある．

「専門性の壁」の正体は，これまで考察したように，専門分野の「概念体系」を知らない人が，専門分野に関する「認識の枠組み」を持たないため，知識を求めるための「問い」を立てにくいことであった．そこで本研究では，専門分野の「概念体系」を明示的に提示する方略を持った情報検索技術を開発し，これに原子力安全分野の概念体系を挿入して「専門的知識へのゲートウェイ」として社会に実装する．

2. 材料と方法

本研究では分野特化型の情報検索システム「Noocle システム」を開発し，これに原子力安全分野の「オントロジー」を入力して，インターネット上で検索サービス実装の公開実験をする．「Noocle システム」はインターネット上のドキュメントを検索する，いわゆる Web 検索エンジンである．「オントロジー」とは，専門分野の索引語に簡潔な説明をつけたものを階層的に分類し，索引語の相互の関係を属性付きのリンクで表現した，「概念体系」のデータのことである．文献 [5]によれば情報工学の分野でオントロジーは「概念化の明示的な規約」などと定義される．この「Noocle システム」の特徴は，検索質問の入力時にユーザにオントロジーを提示することで，対象とする専門分野の「概念体系」の総覧をユーザが俯瞰的に見渡して探索し，疑問を言語化できるように設計されている点である．この章では，開発される「Noocle システム」と実験で公開される「Noocle 検索サービス」とを順に説明する．

2.1. 開発される Noocle システム

「Noocle システム」は，オントロジーを専門家から抽出して形式化するためのオントロジーエディタ「OntStar」，検索対象のドキュメントのインデクスを維持管理するためのコアシステム「Noocle-Core」，検索質問を解釈してユーザにサービスを提供するための検索イ

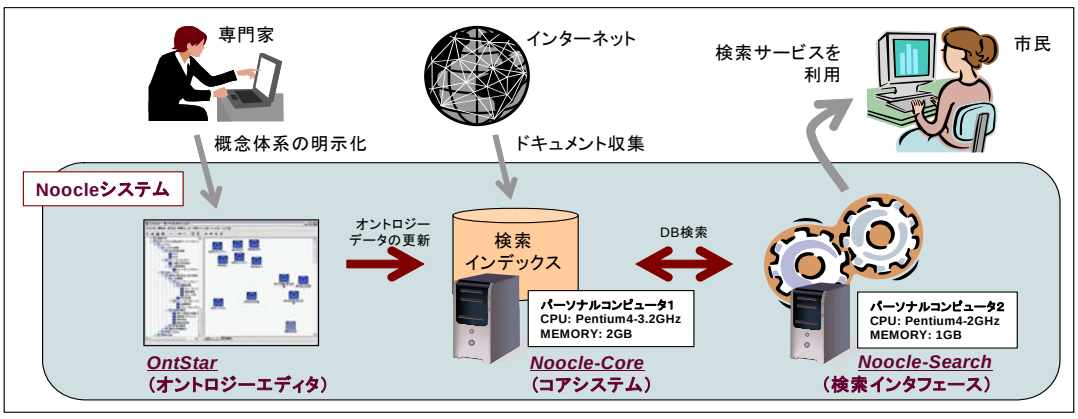

Fig. 1　　Noocle システムの概要

ンタフェース「Noocle-Search」の 3 つのソフトウェアモジュールから構成される．構成の概要を Fig. 1 に示す．以下ではこれらを順に説明する．

(1) オントロジーエディタ「OntStar」

OntStar は専門分野の概念体系を明示的に記述するために筆者らが開発したソフトウェアツールである．Fig. 2 に示すユーザインタフェース画面を持ち，概念の名辞，説明，及び概念間の関係の入力を受けてオントロジーとしてまとめ，Noocle-Core が解釈できるデータファイルを出力する．名辞は表記ゆれや多言語化に対応できるように一つの概念に複数を指定できる．概念間の関係は任意に定義して用いることができるが，上位概念から下位概念への関係はあらかじめ用意されている．これはトピックの分類を想定した関係性であり，「類と種差」のような厳格な関係よりも緩い関係である．一つの概念に複数の上位概念を指定することが許されているが，それぞれの概念は自分の下位概念を上位概念に持つことは許されていない．従って記述される概念の基本的な構造は非循環有向グラフとなる．このソフトウェアは論文執筆時点でベクター（株）のホームページから無料で公開されており，2007 年 9 月 28 日の時点で 428 回ダウンロードされている．

(2) コアシステム「Noocle-Core」

Noocle-Core は検索対象のドキュメントを収集し，検索インデクスを付与して保守するソフトウェアシステムである．具体的には MySQL AB 社製リレーショナルデータベースエンジンを中心として，筆者らが開発した「クローラ」と「インデクサ」が実装されている．性能としては，約 10 万のドキュメントの検索インデクスを保守する場合には 10 日以内のサイクルで情報を最新に保ち続けることができる．処理は完全自動化されている．

Noocle-Core の「クローラ」はインターネット上のリンクをたどって HTML ドキュメント及び PDF ドキュメントを収集するプログラムである．クローラは管理者によって指定される「検索対象サイト」のドキュメントを網羅的に収集してインデクサの処理をとおし，データベースに登録する．このクローラには，対象 Web サーバの負荷を調節する機能，前回訪問時から変更されていないドキュメントの処理を省略する機能，「norobot」タグ等による著作権者の指示を解釈する機能，CGI と呼ばれるプログラムで提供されるコンテンツに対応する機能，などが実装されている．「検索対象サイト」は管理者が管理用インタフェースから任意の Web サーバの任意のディレクトリ以下を条件付きで指定する．検索対象サイトを限定することにより，汎用目的の Web 検索サービスに比べて検索結果の品質が向上し，また安価な計算機資源でも実用的な検索サービスを提供することができるようになる．

「インデクサ」とはドキュメントに検索のための索引情報を付与するプログラムである．Noocle-Core のインデクサは OntStar で与えられるオントロジーのデータファイルを解釈し，概念の名辞をドキュメントから検出してインデクスを作成する．インデクスはオントロジーの各概念に対するドキュメントの関連性をベクトル化した「オントロジーベクトル」[6]の形式で表現されてデータベースに登録される．全システムをオンラインにしたままオントロジーデータを差し替えられるよう設計されており，約 2 千の概念からなるオントロジーを更新した場合，約 10 万のドキュメントの検索インデクスを約 10 時間で再構築することができる．

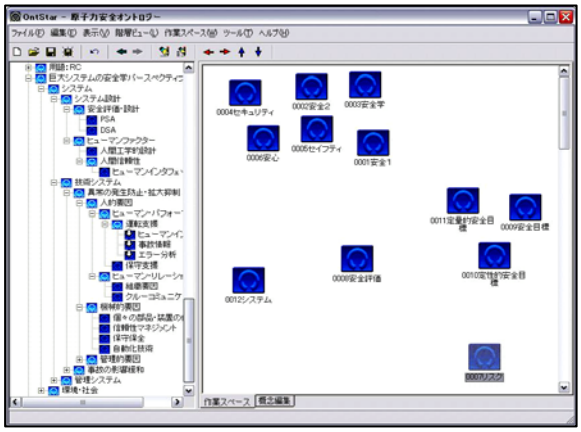

Fig. 2　　OntStar のユーザインタフェース

(3) 検索インタフェース「Noocle-Search」

Noocle-Search は，インターネットを介してユーザに検索質問の入力支援機能と検索機能とを提供する，筆者らが開発した Web 通信処理ソフトウェアである．検索質問の入力支援機能として，Fig. 3 のように概念体系の総覧を提示する設計になっている点が特徴的である．検索機能としては，ユーザから受信した検索質問を Noocle-Core に送り，生成される検索結果セットを Fig. 4 のように整形してユーザの Web ブラウザに表示する．

Fig. 3 の「トップフレーム」はユーザが検索質問を入力して Noocle-Search に対して検索処理を要求するためのユーザインタフェースである．オントロジー中の専門的概念と検索キーワードの 2 つの項目について，それぞれ複数入力することができる．ここで検索対象の専門的概念は「ボトムフレーム」の操作で入力することができる． Noocle-Search は Noocle-Core に対して「ブーリアンモデル」と呼ばれる方法でキーワードによる検索演算を行わせ，得られたドキュメントのサブセットに対して「ベクトルモデル」と呼ばれる方法で概念による検索演算を行わせる．ベクトルモデルの演算では「オントロジーベ

クトルモデル」[6]によって，オントロジー上で検索対象概念に強く関連する概念の名辞を多く含むドキュメントほど適合度が高いと評価する適合度計算をする．仕様として，検索キーワードが指定されなかった場合には収録されている全てのドキュメントに対して適合度を評価し，適合度上位のドキュメントを出力する．検索対象概念が指定されなかった場合には便宜的にオントロジーのトップノードに位置している概念を検索対象概念として外挿して上述の検索を実行する．両方とも指定されなかった場合には空集合を出力する．Noocle システムのキーワード検索機能は，用意されるオントロジーが不完全である場合を想定し，未収録の概念の名辞を検索条件に加えられるように装備された．しかし，実際上はこれを利用して一般のキーワード検索のように任意の検索質問を表現することができる．

Fig. 3 の「ボトムフレーム」はオントロジーを提示し，その構造をユーザに探索させて検索対象概念を指定させるユーザインタフェースである．この中の「ナビゲーションフレーム」と「リストフレーム」は相互に連動して動作する．「ナビゲーションフレーム」では OntStar によって非循環有向グラフの構造で構築されたオントロジーを木構造に展開して提示する．「リストフレーム」はすべての概念の詳細を列挙して提示する．ユーザは関心のある概念を見つけたら「検索対象に追加」ボタンを押して「トップフレーム」の「検索対象概念」欄に入力することができる．オントロジーデータを差し替えた場合，Noocle-Search は「ボトムフレーム」の表示を制御する HTML ファイルを自動的に再構築する．

Noocle システムが検索質問を処理して検索結果を表示するためにかかる処理時間は，Noocle-Core に約 10 万のドキュメントが登録されている条件では，検索対象概念のみを指定したときに約 3 秒，キーワード検索機能を併用したときに約 10 秒である．

2.2. 公開される Noocle 検索サービス

分野特化型情報検索技術の社会的実装の実験として，「Noocle 検索サービス」をインターネット上に公開する．これは開発した Noocle システムに「原子力安全オントロジー」を導入し，原子力関連の Web ドキュメントのインデクスを収集させたものである．「原子力安全オントロジー」について，筆者らは原子力安全に関する 2,133 概念を収集して，一般的な安全学の概念構造に従って整理した [7]．「検索対象サイト」については，政府系，企業系，市民団体系などのサイトを調査して約 130 を指定した．指定されたサイトの一部を Table 1 に示す．これらから自動的に収集されるドキュメント数は実験期間中に 10 万

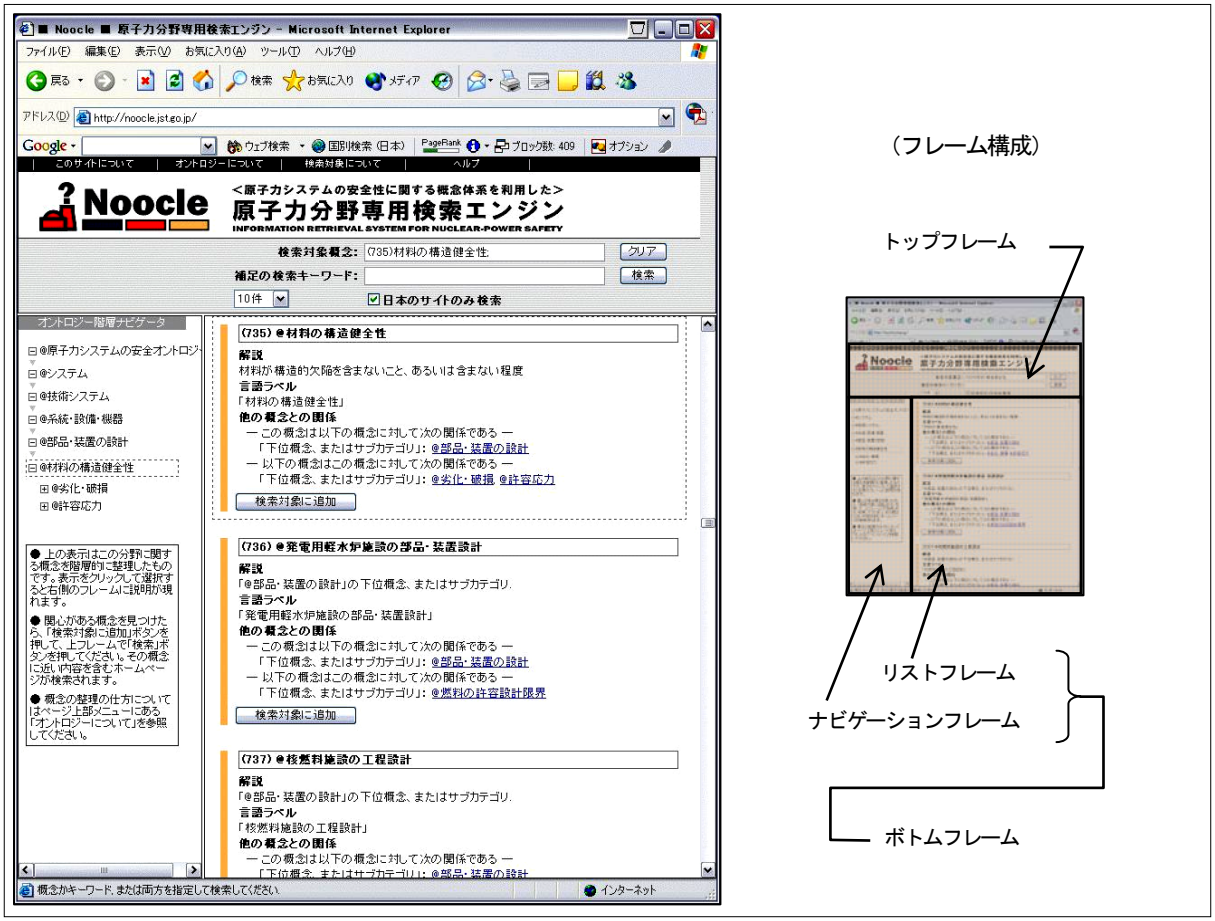

Fig. 3　　Noocle-Search 検索ユーザインタフェース

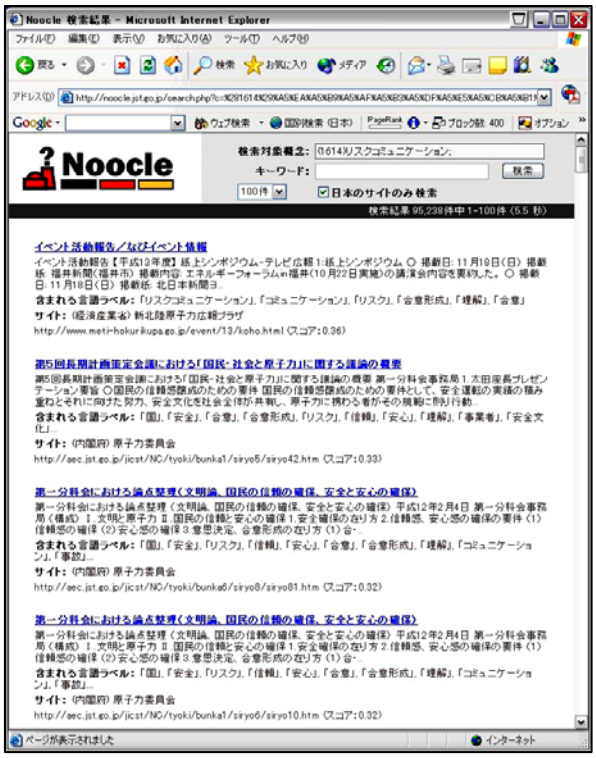

Fig. 4　　Noocle-Search 検索結果表示

付近で推移した．ピーク時には 13 万に達した．

Noocle 検索サービスのインターネット上での公開は利用制限なしに不特定多数の市民が利用できる形で行う．公開期間は 2004 年 12 月から 2006 年 3 月までの約 16 ヶ月間である．「原子力安全オントロジー」の構築作業には約 2 年，原子力関連サイトのリストアップ作業には約 3 週間が費やされた．

3.　実験の結果

Noocle 検索サービスの公開実験の経緯を Table 2 に示す．公開実験期間中の Web アクセスログを Table 3 に示す．ユーザが送信した検索質問の形式別の集計を Table 4 に示す．この検索質問で使用されたオントロジー中の概念の頻度上位を Table 5 に示す．同様に検索キーワードの上位を Table 6 に示す．Noocle 検索サービスがインターネット掲示板で言及された例を Fig. 5 に示す．原子力に関して副次的な社会的問題の発生を避けるために，ユーザの素性に関しては系統的な情報収集を行っていない．

4.　考察

4.1.　社会的実装の成果について

公開実験期間中に，Noocle 検索サービスに対して不特定多数のインターネットユーザから 13 万ヒットを超えるアクセスが行われ，さらに当サービスで多数の検索質問が観測された．従って実験期間中は開発した社会技術が社会に実装されて実際に機能したと結論づけることができる．これによって本研究の第一の目的は達成されたといえる．　Fig. 5 のインターネット掲示板への書き込みからは，本システムから社会への専門的知識の導入の事実があったことが示唆される．さらに次に示すように，受信した検索質問の分析結果は，Noocle 検索サービスが非専門家によって「専門的知識へのゲートウェイ」として利用されたことを示すものであった．

この仮説の検証にあたっては，ある「ことば」に対する大規模検索サイトでのヒット数が，そのことばの一般性または非専門性のおおまかな指標になる，と仮定する．例えば，2008 年 2 月現在，商用の大規模検索サイトの一つを使って「"環境"」という一般的なことばの文字列を検索すると 49,700,000 件のヒットがあり，一方で「"非常用炉心冷却装置"」という専門的なことばの文字列を検索すると 2,610 件のヒットがある．

公開実験中に受信された「検索対象概念」(Table 5)の名辞について，同商用サイトでの「ヒット数」の受信毎の平均は 18,200,000 件 ($\sigma^2=2.44\times10^{15}$)，「検索キーワー

Table 1　収集ドキュメント数上位の検索対象サイト

検索のサイト（ドキュメント数順）	ドキュメント数
(政府) 原子力委員会	18,325
(株) 日本原燃	8,046
(政府) 原子力安全委員会	7,188
(独) JST 原子力百科事典「ATOMICA」	5,693
(独) 放射線医学総合研究所	5,473
(独) 日本原子力研究所	4,512
(株) 東京電力「原子力への取り組み」	4,036
(社) 日本原子力学会	2,410
(財) 放射線影響研究所	2,368
（政府）文部科学省 原子力安全課 環境防災 N ネット	2,095
（財）原子力安全研究協会 緊急被ばく医療情報ネットワーク	1,998
（政府）文部科学省 原子力・放射線の安全確保	1,954
(独) 原子力安全基盤機構	1,929
(財) 放射線利用振興協会	1,850
(福井県) 原子力安全対策課	1,633
(NGO) 原子力資料情報室	1,629
その他（全 133 検索対象サイト，全 94,801 ドキュメント）	

※2006 年 2 月 27 日現在．

※NGO，市民団体や個人のサイトも多数を収集対象としたが，それらの多くはこのドキュメント数上位に入らなかった．

Table 2　サービス公開実験の経緯

2004 年 12 月上旬	公開実験開始
2004 年 12 月中旬	Yahoo!カテゴリ「原子力」に掲載
2005 年 1 月 19 日	受信した検索質問の記録開始
2005 年 4 月 20 日	拠点移動のためサービス停止
2005 年 5 月 12 日	サービス再開
2006 年 3 月 31 日	公開実験終了

Table 3　Noocle-Search のアクセスログ

月	ヒット数	ページビュー数
2004 年 12 月	11,417	3,683
2005 年 01 月	11,876	3,731
2005 年 02 月	7,377	2,249
2005 年 03 月	4,978	1,594
2005 年 04 月	3,845	1,240
2005 年 05 月	4,587	1,279
2005 年 06 月	10,503	3,240
2005 年 07 月	10,912	3,496
2005 年 08 月	18,722	2,560
2005 年 09 月	9,441	3,217
2005 年 10 月	7,063	2,234
2005 年 11 月	6,645	2,101
2005 年 12 月	6,467	1,953
2006 年 01 月	7,701	2,266
2006 年 02 月	5,116	1,621
2006 年 03 月	4,436	1,507
16 ヶ月間合計	131,086	37,971

※　Web サーバ用ログ解析ソフト Webalizer で集計.
※　「ヒット数」は HTTP リクエストの全数（ローカルキャッシュ時刻の問い合わせや通信エラーを含む）.「ページビュー数」は HTML ドキュメントの実送信数.

ド」(Table 6)については平均 3,390,000 件 ($\sigma^2=7.25\times10^{14}$) であった．また，用意した「原子力安全オントロジー」の全概念の個別の名辞に対して，「ヒット数」の平均は 5,320,000 件 ($\sigma^2=8.46\times10^{15}$) であった．ただしいずれも 2 バイト以下の文字列は除いて評価した.

「検索対象概念」は，ユーザがオントロジーを利用して検索質問として指定したものであるが，この専門性は他の 2 つよりも際だって低いように見える．実際に，分散分析法と多重比較法を使用して統計的有意差の検定を行ったところ，「原子力安全オントロジー」と「検索キーワード」の「ヒット数」の間には有意差が認められないが，「検索対象概念」の「ヒット数」はこれらに比べて有意に高かった ($p<0.05$)．ユーザが「検索対象概念」を探索する時には，Noocle システムは常にその下位概念も操作画面に表示する設計になっているため，専門的な疑問を持ったユーザが不本意に粒度の大きい概念を検索したとは考えづらい.

「原子力安全オントロジー」の語彙の専門性が専門家の使う語彙の専門性の基準であると仮定すれば，キーワードを利用した検索は専門的な検索質問を発行するために利用されるのに対して，オントロジーを利用した検索は，比較的に一般性が高い検索質問を発行するために利用される傾向があったと結論づけられる．従って，検索質問の専門性とそのユーザの専門性とが一致すると仮定すると，「Noocle 検索サービス」の概念体系を提示する機能が，非専門家によって「専門的知識へのゲートウェイ」として利用された，という仮説が支持される.

なお，Noocle システムは，検索対象として一般性の高い概念を与えられた場合にも保持するオントロジーを利用して関連性の高い専門的なドキュメントを提供する「検索質問拡張」の能力を有している.

Table 4　検索質問が受信された回数の形式別集計

期間	概念のみ	KW のみ	概念＋KW	入力なし	小計
2005 年 01 月	12	177	12	89	290
2005 年 02 月	20	192	21	30	263
2005 年 03 月	12	94	10	23	139
2005 年 04 月	50	98	10	25	183
2005 年 05 月	34	112	74	51	271
2005 年 06 月	21	257	23	75	376
2005 年 07 月	14	209	8	33	264
2005 年 08 月	24	182	17	20	243
2005 年 09 月	12	167	16	23	218
2005 年 10 月	12	155	2	13	182
2005 年 11 月	10	90	9	112	221
2005 年 12 月	23	60	10	39	132
2006 年 01 月	11	154	12	17	194
2006 年 02 月	13	67	8	40	128
2006 年 03 月	7	61	11	57	136
15 ヶ月間合計	275	2075	243	647	3,240

※　「概念」はオントロジー中の概念を指定した検索,「KW」は任意のキーワードを指定した検索.
※　2005 年 1 月 19 日から 2006 年 3 月 31 日まで集計.

4.2. 技術的な成果について

技術的な成果として，ここではサービス提供のコストの低さ，技術の完成度，そして技術の新規性を議論する.

人手が介在する検索サービスを提供する場合に現実的に問題となるのは，便益に対して「そのコストが許容できるか」という点である．Noocle システムを用いて検索サービスを提供するコストについては，次に分析するように初期コストも運用コストも非常に低く抑えることができると考えられる．初期コストについて，設備コストはパーソナルコンピュータ 2 台分であった．人的コストについては主にオントロジーの構築のコストであるが，Noocle システムは任意のキーワードで検索することもできる仕様になっており，これによってオントロジーの内容が十分ではない段階からでもサービスを提供するこ

Table 5　受信された検索対象概念の頻度上位

基本概念	(67)	自然放射線	(43)
安全	(43)		
原子力システムの安全オントロジー			(33)
リスク	(22)	ハザード	(22)
Moody の臨界流モデル	(21)	システム	(15)
安心	(12)	放射性物質	(12)
電磁波	(12)	健康リスク	(11)
外部電源喪失	(10)	γ線	(9)
環境	(9)	安全保護系の作動ロジック	(8)
労働衛生管理	(8)	破断口	(6)
セキュリティ	(6)	コントロールセンタ	(5)
放射能	(5)	PSA パラメータ	(5)
核燃料物質	(5)		
		【のべ使用回数：664；異なり数：192】	

※　5 回以上使用された概念について表示．カッコ内は使用された回数．「原子力システムの安全オントロジー」は入力画面の最上部に表示されるトップノード．

Table 6　受信された検索キーワードの頻度上位

地球への影響	(36)	安全解析	(30)
放射性物質	(29)	コバルト 60	(22)
危険性	(19)	中レベル放射性廃棄物	(17)
コバルト	(16)	放射線	(16)
臨界流	(15)	γ線	(15)
外部電源喪失	(15)	崩壊熱	(14)
最高使用温度	(14)	テロでの年間死亡数	(14)
低レベル放射性廃棄物	(13)	ウラン	(13)
PSA	(13)	減衰係数	(12)
日本原子力技術協会	(12)	原子炉	(12)
減幅比	(11)	キセノン	(10)
非確率的影響	(10)	操作時間	(10)
安全性	(10)		
		【のべ使用回数：2318；異なり数：1024】	

※　文字化けしたデータを除いて，検索に 10 回以上使用されたキーワードについて表示．カッコ内は使用された回数．最下行の集計は文字化けしたデータを含む．

とができる．具体的に想定すると，「OntStar」を使用して対象分野についての教科書的な書籍の索引欄から索引語を抽出してオントロジーを構築するなどすれば，必要な準備期間は数週間程度に収まると見積もられる．さらに過去のオントロジーを再利用できる場合はより短い準備期間でサービスを開始できると考えられる．運用コストについて，設備コストは電気料金と通信回線の維持費のみである．人的コストについては，オントロジーと検索対象サイトの拡充の努力によるので高くも低くもなり得るが，いずれにせよこれらの操作は 2 章で説明したように簡便に行えるよう設計されているので，誰でも運用することができる．従って実際のサービス提供のコストは非常に低く抑えることができると考えられる．

ソフトウェアの完成度については，Noocle システムは公開実験期間の 16 ヶ月間にわたり安定してサービスを提供し続け，必要なメンテナンスは検索対象サイトの更新のみであった．従ってその完成度は高い水準に達していると考えられる．

オントロジーを利用した分野特化型情報検索の先行技術としては，Unified Medical Language System (UMLS)を利用して MEDLINE を検索する PubMed などの検索サイトが実用化されている．MEDLINE は 1600 万報以上の医学文献を収録する米国国立医学図書館のデータベースである [8]．UMLS は 200 万語以上の語彙を収録する医学分野で整備されたオントロジーであり [9]，医学の専門家向けに UMLS を利用して検索質問拡張を行う分野特化型情報検索サービスが，この PubMed 以外にも複数開発されている．「検索質問拡張」とは，入力された検索キーワードと共にそれに関連する別の適切なキーワードも同時に検索対象とする手法である．さらに非専門家を対象とし，UMLS を利用して検索サービスを提供する研究としては，非専門家向け検索技術「MedicoPort」[10]の研究開発などが行われている．MedicoPort は，例えば非専門家が一般語で「wet bed」（おねしょする）と入力すると「nocturnal enuresis」（夜尿症）という専門用語による検索結果を提示するとされている．この研究では医学系の Web ページを選択的に収集する方法としてページのリンクを評価している点で，本研究と異なっており興味深い．ただし概念体系をユーザに提示する機能はない．これらの技術に対する本研究の新規性は，専門分野のオントロジーを非専門家に提示する「認識の枠組み」として利用したこと，そして「専門的知識へのゲートウェイ」を指向して分野特化型情報検索技術を実装する場合，Noocle システムのソフトウェア設計によって，有効で安定した実サービスの提供が現実的なコストで実現可能であることを実証したことである．

4.3. 概念体系が認識にバイアスを与える問題について

本研究では非専門家に専門分野の認識の枠組みを提供する技術を提案しているが，これは次に議論するように認識にバイアスをかける能力があるという点で危険性がある．この対策として任意の概念体系を使用した検索サービスを誰でも構築できるように技術を公開するなどの工夫が必要であると考える．

これまでの研究 [11]で，原子力のエンジニアのコミュニティと原子力撤退派の市民のコミュニティとでは異なる概念体系を持っている可能性が指摘されている．それぞれの思想が発達した社会的背景を考慮すれば，それぞれの概念体系に価値判断に基づく系統的なバイアスが掛かっている可能性がある．また，本研究ではリスク概念に重点をおいたオントロジーを提供したが，このような「リスクとベネフィット」や「リスクと安全」といった二項対立の設定は，原子力技術の認識にバイアスを与える可能性をはらんでいる．具体的には，例えば哲学者のウィ

ナーが文献[12]の中で，原子力などの「リスク・ベネフィット分析」の議論について，これが規制を緩和する方向に誘導的であると指摘している．さらにリスクと安全の対立概念に関しては，「リスクと安全をめぐる論争が，原子力についての人々の議論の範囲をかなりの程度まで形作ってしまい，ほかのすべてのことを排除してしまった」（p. 247）と指摘している．これらの指摘のように，二項対立を設定する認識の構造化は，対立軸に乗らない他の要素を捨象して結論を誘導する能力を持っているといえる．

つまり，概念体系には価値判断が混入する可能性があり，また概念上の二項対立の設定も作為的に行われる可能性があるといえる．このことから「認識の枠組み」として与えられる概念体系は相対化されて批判されていなければ危険であるといえる．この問題に対処する方法の一つは，概念体系の背後にある思想の複数性を認めることであると考えられる．複数の概念体系が提供されることにより，これを利用する非専門家は問いを立てるための認識の枠組みを任意に選択し，または切り替えながら知識を探求することができるようになると考えられる．

Noocle システムのように検索サービスを低いコストで構築できる技術が一般に提供されれば，この要件を満足できる可能性がある．検索サービスの構築環境が一般に公開されることが前提となるが，前節で述べたように低いコストで独自の検索サービスを提供できるならば，様々な視点を持つ専門家にとっての思想の表現手段として検索サイトが構築され，結果として同様の問題領域に関する複数の概念体系が社会に提供される可能性がある．

これを実現するためには，オントロジーデータは既存のものを組み合わせたり手直ししたりして自由に再利用できるように，本研究で提案するような分野特化型の情報検索技術は誰もが思想の表現手段として自由に利用できるように，社会に実装される必要があると考える．

4.4. 科学一般の知識流通に応用する場合について

本研究では分野特化型の情報検索技術を原子力安全分野の知識流通のために利用したが，この技術は科学一般の知識流通のためにも応用することができると考えられる．科学の領域は専門分野が細分化しているので，この場合には次に議論するように個別の科学者コミュニティがそれぞれの検索サービスを提供することが望ましいと考えられる．

ある専門領域である「ことば」が専門用語として用いられる場合には，実際の「意味範囲」がその語の素朴なイメージとは異なっていることがある．例えば原子力工学分野では「安全性」という専門用語は「原子力システムがシステム内部の放射能を閉じこめておく能力の確実性」という独特の意味で用いられる．このように，相当数の専門用語がそれぞれの専門分野で独特の意味で用いられていると思われる．

言語学者のソシュールは，言語を「ラング」と「パロール」に区別してその関係を考察している[13)14)]．「ラング」とは語彙と文法の体系を指す概念であり，本研究で扱っている「概念体系」を包含する概念である．「パロール」とは個別の言語活動を指す概念であり，本研究で扱っている「ドキュメント」を包含する概念である．ソシュールは，言語活動の蓄積が語彙と文法の体系を規定し，語彙と文法の体系が言語活動を規定するとして，ラングとパロールが循環的に発達する構造を指摘している．この指摘によれば，科学者のコミュニティの中で専門用語に固有の意味が割り当てられ，新たな「構成概念」が生産されていくことが説明できる．そして分野特化型の情報検索サービスについては，概念体系と，それによって解釈可能なドキュメントとを対応づけて科学者コミュニティごとに扱うべきであることが推論される．Noocle システムのようにオントロジーと検索対象サイトとを一組にして取り扱うソフトウェア設計は，科学者コミュニティ

掲示板名： 「地震があったので臨時@2ch」
スレッド名： 「【浜岡厨】浜岡原発を何とかスレ 16【追悼】」

◆12at95.4iI ：05/02/01 03:03:28 ID:rlbUij2M
つか、今まで寡聞にして知らずだったけど、原子力関係の用語説明サイトって
感動的に充実してるのね。びっくりした。

原子力百科事典 ATOMICA/原子力図書館げんしろう
http://mext-atm.jst.go.jp/atomica/
■ Noocle ■ 原子力分野専用検索エンジン
http://noocle.jst.go.jp/

↑の下側のやつによると「便益あたりリスク」などという、さらにややこしいことになりそうな概念もある模様。
たしかに、便益でかかったら、多少リスクでかくても他でカバーできるかもしれないし、
逆にメリットもないのにリスクでかかったら、意味ないもんなぁ。

Fig. 5 実験中の Noocle 検索サービスがインターネット掲示板で言及された例

ごとの情報検索サービスを提供するための要件を満たしているといえる．

科学者のコミュニティの切り分けに関して，藤垣は科学技術社会論に立脚して「ジャーナル共同体」という結節点を提案している[15]．藤垣はピアレビューを行っている科学者の知識生産活動がジャーナルの差別化とジャーナル採録基準への適合化の双方を最大にすることを指向していることを指摘して，「ジャーナル共同体」が個別分野のディシプリンを確立して維持する主体に一致するとしている．従って自然科学や科学技術に関する「専門的知識へのゲートウェイ」を提供する場合には，この主体の一部はジャーナルを刊行する各学会によって担われるべきであると考えられる．

4.5. 開発した社会技術の今後の展望

前節では，ピアレビューの文化を持つ自然科学系の分野を考察した．しかし他にも，構造化された知識を確立して維持している専門分野ならば，本技術によってその知識へのゲートウェイを提供することが可能であると考えられる．また次に議論するように，そのような社会的実装は有用であると考えられる．このような専門分野としては，人文社会科学系の分野は当然のことながら，各種の業界，思想，宗教，趣味，サブカルチャーなど，さまざまな分野が該当する．

2007 年現在のインターネットの情報量は，静的な Web ページ数だけでも数百億ページといわれ，さらに急速に増加しつつあるといわれる．このような状況は「情報爆発」と形容される．論文執筆時点でこれらの情報を利用するための主な手段はキーワード検索技術である．キーワード検索技術は，原理的にはキーワードの組み合わせによってインターネット上の全ての知識にアクセスできることから，非常に有用な技術であるといえる．しかし人が使える語彙は限定されているので，実際に問題解決のためにインターネットを利用しようとする場面では，問いを表現するためのキーワードが未知であるような困難な状況にであうことが予想される．このような，何を問えばいいのかが分からないために問題を把握できず，そのために結局何を問えばいいのか分からない，という悪循環は「メノンのパラドックス」と呼ばれる[16]．

人の具体的な語彙数について，教養のある英語の母語話者が持つ「受容語彙」は約 10 万語とされるが，「使用語彙」はわずか 1 万語から 2 万語とされる[17]．受容語彙とは他人が使えば理解できるが本人が使うことはない語彙，使用語彙とは本人が理解していて使うこともできる語彙のことである．いずれの言語圏に属する人でもこの語彙数に大差はないと思われる．例えば国語辞典「広辞苑第五版」の総項目数は 23 万項目あり，さらに国語辞典に収録されない固有名詞や専門用語，業界用語等が膨大に存在するであろうことを考えると，我々が使用できるたかだか 2 万語の語彙は，インターネット上の全ての知識の索引としてはまったく貧弱である．この問題は，人々が自分の専門以外の構成概念を知らないために情報要求を言語化できないという，序論で述べた問題を包含していると考えられる．使用できる語彙や概念が制約されている人間の問題解決をエンジニアリングの立場から支援するためには，人の認識の原理に立ち返った「構造主義」的な検索技術を提供することが必要であると考えられる．

Noocle 検索サービスのように概念体系を提示する分野特化型情報検索サービスが，各種の学問分野，業界，思想，宗教，趣味，サブカルチャーなどのさまざまな分野で構築されれば，個人が持っている語彙の限界を乗り超えて問いを立てられるようになる可能性がある．このようなサービスは問題解決の一般的手段として社会的に有用であるといえる．Web 広告ビジネスの収益モデルは成熟しつつあるので，そのようなポータルサイトが誰にでも低コストで構築できるならば，様々な分野の専門的知識を持つ人にとってサービス提供の動機付けは十分にあると考えられる．

多くの分野に関する多くの情報検索サービスが分散的に提供されるとすると，それらのサービスを検索するようなメタな検索サービスを提供することも可能になる．これによって，インターネット上に散在する知識の全般的な構造化を実現することも可能になると考えられる．このように，様々な専門分野で Noocle 検索サービスのような「オントロジーを利用した分野特化型の情報検索技術」が提供されれば，インターネット上の知識流通と人々の問題解決のあり方を改善できる可能性がある．

5. 結論

本論文では，非専門家による専門的知識の探索を支援する「専門的知識へのゲートウェイ」として，社会技術「オントロジーを利用した分野特化型の情報検索技術」を提案し，その社会的実装の実験について報告した．さらにこの技術の能力と課題，展望について考察を加えた．実験は知識流通の不全がもたらす問題が顕在化している原子力分野を対象とした．実験期間中に不特定多数のユーザからの実際的な利用があったことが確認されており，社会的実装の目標は達成されたといえる．この「専門的知識へのゲートウェイ」の社会技術は，他の様々な分野の知識流通にも利用できると考えられる．

参考文献

1) 徳永健伸 (1999)『情報検索と言語処理』東京大学出版会.
2) Taylor, R. S. (1967). Question-Negotiation and Information - Seeking in Libraries. *Studies in the Man-System Interface in Libraries*, Report No. 3, Center for the Information Sciences, Lehigh University.
3) 高田明典 (2007)『構造主義がよ～くわかる本―人間と社会を縛る構造を解き明かす』秀和システム.
4) National Research Council. (1996). *Understanding Risk: Informing Decisions in a Democratic Society*. Washington, D.C. : National Academy Press.
5) 溝口理一郎 (2005)『オントロジー工学』オーム社.
6) 尾暮拓也, 中田圭一, 古田一雄 (2005.11)「コミュニティオントロジーを利用した情報検索」『社会技術研究論文集』3, 102-110.
7) Furuta, K., Ogure, T., and Ujita, H. (2005). Nuclear Safety Ontology - Basis for Sharing Relevant Knowledge among Society. In Arai, T., Yamamoto, S., and Makino K., (Eds.), *Systems and Human Science for Safety, Security and Dependability* (pp. 397-408). Amsterdam: Elsevier.
8) U.S. National Library of Medicine (2007). *MEDLINE Fact Sheet*. http://www.nlm.nih.gov/pubs/factsheets/medline.html [2007, October 15].
9) Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32, D267-D270.
10) Can, A. B. and Baykal, N. (2007). MedicoPort: A medical search engine for all. *Computer Methods and Programs in Biomedicine*, 86, 73-86.
11) 尾暮拓也, 高松悠, 古田一雄 (2004, 10)「コミュニティを超えた知識共有のための原子力安全オントロジー設計方法」『社会技術研究論文集』2, 389-398.
12) Winner, L. (2000)『鯨と原子炉―技術の限界を求めて』(吉岡斉, 若松征男訳) 紀伊國屋書店 (1986).
13) Saussure, F. (1972) 『一般言語学講義』 (小林英夫 訳) 岩波書店 (1949).
14) 加賀野井秀一 (2004) 『知の教科書 ソシュール』講談社.
15) 藤垣裕子 (2003)『専門知と公共性－科学技術社会論の構築へ向けて』東京大学出版会.
16) Polanyi, M. (1967) 『暗黙知の次元』 (高橋勇夫 訳) ちくま学芸文庫 (2005).
17) 田中春美, 樋口時弘, 家村睦夫, 五十嵐康男, 下宮忠雄, 田中幸子 (1994) 『入門ことばの科学』大修館書店.

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Ⅰ「安全性に関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振興機構の事業）の研究として行われた. 本研究は文系と理系の知見を統合して社会問題の解決に取り組むという社会技術の方法論を実践した. この事業で交流した各分野の研究者, および「原子力安全Ⅰ研究サブグループ」のプロジェクト「Design Initiative for Risk Aware Society (DIRAS)」に関わられた皆様に感謝いたします. この論文にはレビュアーからの貢献も含まれています. レビューでは「専門的知識へのゲートウェイ」の実装の別アプローチとして, 既存の商用検索サイトの Web API を利用して検索質問の入力支援を行うという, 発展的な提案を頂きました.

SOCIAL IMPLEMENTATION OF ONTOLOGY-BASED DOMAIN-ORIENTED INFORMATION RETRIEVAL

Takuya OGURE [1] and Kazuo FURUTA[2]

[1]Ph.D. (Eng.) National Institute of Advanced Industrial Science and Technology (E-mail: ogure.takuya@aist.go.jp)
[2]Ph. D. (Eng.) Professor, The University of Tokyo, QUEST (E-mail: furuta@q.t.u-tokyo.ac.jp)

Domain knowledge, which experts have accumulated locally, should be shared with other non-specialists more broadly to aid their *decision makin*g and *problem solving*. One of the largest social issues due to inadequate *knowledge sharing* can be observed, however, in a matter of “public acceptance of nuclear power”, where people can hardly participate in its technical discussion beyond impeding expertise. In order to mitigate the difficulty, we developed an information retrieval service based on an ontological technique. This new service was implemented into society and served there to make it easier to search nuclear-related contents on the Internet. This technology can be applied to any other matters to aid people in reaching specialized domain knowledge.

Key Words: *Information Retrieval, Ontology, Nuclear Safety, NOOCLE, Gateway for Domain Knowledge*