

情報科学技術シソーラスの規格用語への拡張

小川 清, 加藤 輝政, 真鍋 孝顕, 松下 聖一(名古屋市工業研究所)

1 書籍へのシソーラスの付与

書誌事項としてのシソーラス

書籍、マニュアルなどに、書誌事項として、目次、抄録、索引などを付与する意義は広く認められている。しかし、これらの道具だけでは、書籍中の用語の意味が必ずしも明確にならない場合もある。専門用語などが多く使われている書籍の場合、書籍の最後に用語集などを付けることがある。

しかし、複数の学問分野にまたがる書籍の場合には、これらの用語集だけでは十分でない。複数の学問分野にまたがる書籍へのシソーラスの付与を提案し、簡易なシソーラスを作成してきた。(1)

書籍に付与するシソーラスの設計

書籍に付与するシソーラスの設計は、書籍の構成の設計と並列に行うこととした。目次の概念構成と、シソーラスの概念構成を同一過程で行うことにより、シソーラス自体が書籍の理解を助ける道具として役立つような設計を行った。

当初は9つの概念で出発したが、基本的な概念を十進分類法のように10の基本概念に整理し直した。基本概念を理解するための用語を列挙し、本文の理解を助けるものとした。この10という数字に意味はないが、十進分類が広く普及してきたことと、人間が一次元で把握可能な概念の一つの上限が10であるという仮定に基づいている。

その後、内容を改訂したものを毎年作成したが、それらはシソーラスの改訂と並列に行った。内容の改訂はシソーラスの改訂に反映し、シソーラスの改訂は内容の改訂に反映することにより、スパイラル的に内容の改善をもたらすことができた。(2)

2 データベース検索技術シソーラス

シソーラスの目的

データベース検索技術シソーラスの具体的な目的は、データベース検索において、

必要な技術が認識されておらず、データベース検索に必要な技術とは何かを抽出することに目的があった。

特に、商品名と物質名とが一对一对で扱えるような単純な用語構造の学問分野と、商品名・サービス名が、時間的・場所的に複雑に変化しながら提供される情報サービスとを混同しないことを一つの出発点とした。

人間の補助記憶装置としてのデータベースの活用方法が、もし何かを暗記しなくてはならないとすれば、それはデータベース検索技術に対する誤解に基づいているという前提も明確にした。

データベース検索技術の10の概念

データベース検索技術が、商品名・サービス名とは関係ない技術概念が存在することを明確にするため、情報リテラシー、情報システム、ドキュメンテーション、データベース、情報検索、知的財産権、通信、コンピュータ、英語の9つの概念を提案した。さらに、集合、情報、言語、計算、伝達、文献、権利、体系、記録、検索の10の抽象的な概念に再編成した。

3 シソーラスの拡張

複数の用語集の結合

データベース検索技術シソーラスを、情報科学技術シソーラスへ拡張を行った。

一つの書籍におけるシソーラスを編む場合でも、複数の用語集からその上下関係を引用したり、編集したりする場合がある。

複数の用語集の結合は、3つのパターンが考えられる。

- 1 同じ専門分野の辞書の結合。
- 2 異なる分野の辞書の結合。
- 3 専門分野辞書と一般用語辞書の結合。

ここでは同じ分野のシソーラスの結合として、データベース検索技術シソーラスを元に、シソーラスの拡張を行い、情報科学技術シソーラスを作成した。

複数辞書の結合

同じ分野の専門用語辞書

太線化（編集方針の摺り合わせ）

異なる分野の専門用語辞書

X字化（共通部分の確定）

専門用語辞書と一般用語辞書

L字化（基本用語の意味の類推）

同一分野の辞書の結合

同一専門分野の用語同士の結合は、それぞれの目的と傾向を分析することにより、結合はかなり容易である。最初に作成したシソーラスに、岩波「情報科学技術辞典」に付与されている「用語の木」を併合した過程がこれに相当する。

基本的な方針がうまく摺り合わせれる場合には、2つの辞書を足し合わせるだけで、よく、この場合を「太字化」と分類する。

2つの辞書の方針が異なるが、表記方法、項目などが異なるだけで、意味的な方針がうまく組み合わせれる場合には、「より対線化」と分類する。

方針がかなり異なり、そのままではうまく結合できない場合を、「縫い合わせ」と分類する。

ここでの作業は、より対線化であった。

情報科学技術シソーラス

岩波「情報科学技術辞典」に付与されている「用語の木」のすべてと、この用語の木に記述されていない解説中の語の一部を付加して、元のシソーラスの体系が壊れないように配慮した。「用語の木」には、本文中の専門用語がすべて網羅されている訳ではないので、必要に応じ、本文からも用語をピックアップした。

日本語のシソーラス

日本語のシソーラスにおいては、日本語の表記の種類は、「ひらがな」、「カタカナ」、「roma字」、「漢字」表記の4種類必要がある。しかも、ひらがな、カタカナ、roma字、漢字の表記が1種類とは限らないため、それらの表記の集合定義

が煩雑になる。

日本語

ひらがな

カタカナ

漢字

roma字

英語

専門用語

インターネットの検索システムなどにおいて、目的の用語に到達するためには、専門用語のうち、固有名詞を用いることが重要である。なぜならば、一般名詞ではヒット件数が多くなり、固有名詞による限定が有効であるからである。

しかし、固有名詞を知っていることが、その専門分野の知識があることではないこと、その専門分野の商品名、サービス名を知っていることがその専門分野の知識があることを代表しないことは自明のことである。

固有名詞

固有名詞は、基本的にその母国語で表記されるべきである。その英語表記、日本語表記を並記してあると便利な場合がある。そこで、固有名詞は母国語による表記、英語による表記、日本語による表記を並記するものとすると同時に、固有名詞であるための識別記号を付すことにし、当面「」で囲むことにした。

固有名詞

人名

組織名

国名

会社名

商品名・サービス名

登録商標

地名

人名

日本の人名では、コンピュータの基本的なコードにおいて、現存するすべての人名を表記できるコードが存在していないことが、最大の問題である。

こういった事態は、他の主要言語では存在していないことであり、日本語情報処理が、他の言語に比べて、大きく遅れていることの象徴として知られている。

インターネットの国際化が必要になっている今でも、日本語の人名表記をどのように扱うか、日本から積極的な提案がされていないことが、インターネットの国際化を遅らせている要因ともなっている。

組織名

組織名は、しばしば略号を持っている場合がある。そのため、略号とフルスペルとを列記する必要がある。企業名などにおいては、一つの国における企業名が、他国では別の企業が使用している場合に、企業名の国による違いをメモしておく必要がある。その関係で、日本の企業で、海外ではブランド名を企業名に用いている企業もある。

商品名・サービス名

商品名・サービス名は、特定の国において商標として登録されている場合があり、どこの国での登録商標かをメモしてある方が便利である。また、別の国においては、すでに他の商品が登録されている場合には、同じ商品・サービスで国によって名称が異なる場合がある。

4 異なる分野の辞書の結合

異なる分野の辞書の結合の場合は、その共通部分を、同一専門分野の辞書の結合方式に基づいて行う。共通部分の辞書の方針が異なる場合には、より基本的な考え方を取っているものを優先した。

図書館用語の挿入

図書館は、ドキュメンテーションと情報サービスの重要な役割をになってきたため、紙を用いた情報処理技術に関する用語

が豊富であり、異なる分野とはいっても、お互いの穴を埋めるような用語の補充を行うことができた。

言語学用語の挿入

言語学の辞書からの用語の補充では、コンピュータサイエンスの人工言語のための理論的な概念の補充と、自然言語のための理論的な概念の補充ができ、図書館関係の用語の場合と類似的な穴埋めが可能となった。

日常用語の挿入

しかし、日常用語と専門用語をどのように結合させるかは、専門用語の体系、および教育において、知識の構造との相似性を発見する過程としても有効であると考え試作した。

具体的には「角川類語辞典」の概念構成の用語を単純に取り込み、その構造を、従来の構造の指針に基づき組み替え、併合した。

しかし、角川類語辞書の見出し語は、グループ分けできるものが少なく、並列的な概念のものが多いため、そこらかの引用部分だけとなった。

一部の概念においては、それまでに反対語の分類を設けないものもあったが、日常用語の分類から、新たに専門用語の中の対比語、反対語の見直しに役立った。

5 規格用語

情報分野においては、ISOなどの国際規格、JISなどの国内規格、そのほか業界団体規格、企業連合規格などがある。また、特定の企業の仕様がデファクトスタンダードとなることも多く、規格・仕様の用語が、技術用語として定着する場合もある。そのため、情報科学技術シソーラスに規格用語を加えることとした。

規格番号

規格用語を付加するにあたって、規格番号を見出し語に付与することとした。これは、規格番号でも用語が検索できるようにするためである。

漢字（ひらがな、カタカナ、English）、規格番号

BT 上位語

NT 下位語

RT 関連語

表記の統合

2つの辞書で用いられている語の表記を統合するには、2つの方法がある。一つの表記に統一する方法と、2つの表記を見出し語として採用し、一方から他方を指す方法である。

表記があまり違わなければ、一方の表記に他方の表記を併記しても構わない。例えば、「フェース」と「フェイス」は、最初の2文字まで同一であるので、別々に記載して見出し語が増えることを防ぐ意味がある。

規格分類

ISOには、国際規格分類が制定され、規格と規格用語の体系的な理解に役立つものとなっている。

従来の図書分類では、情報技術関連の分類があいまいで、多くの図書館で、情報技術の文献が分散したり、分類方法が不適切な配置になったままとっていた。

このことが、図書館関連諸学において、情報科学に対する体系的な理解の阻害的な要因の一つとも考えられた。

規格用語の結合範囲

規格用語の結合は、ISOのIT(information technology)の分類と、JISのX分類の主要用語をまず対象とした。

しかし、プログラミング言語相互の変換可能性の定義がないなど、同一分野内で複数の用語規定があるなど、規格用語自体が、縫い合わせの必要のある用語体系であることが確認できた。

今後は、ANSI, IEEE, ITUなど共通部分の多いものに範囲を広げていく予定である。

ISOなどのように、規格に著作権が存在し、その規格の内容の普及に対して、必ずしも

好ましい影響を与えていない可能性もあるように思われた。

規格用語全体が見渡せた段階で、特許分類の情報系の用語の結合がその次の課題として設定している。

参考文献

- (1)データベース検索技術者認定試験 傾向と対策、小川清、安藤敏、紀伊国屋書店
- (2)データベース検索技術テキスト、小川清、安藤敏俊、東海サーチャー会
- (3)情報科学技術辞典、岩波書店
- (4)情報科学技術シソーラスの作成、小川清、ドクメンテーションシンポジウム
- (5)データベース検索技術シソーラスの作成、小川清、情報処理学会全国大会
- (6)類語辞典、角川書店