

社会科学情報の概念理論

齊 藤 孝

Concept Theory for Knowledge Organization in Social Science Information

Takashi SAITO

The concept theory provides user' perspectives of subject schema which is concerned with information of digital media. Information scientists have focused on designing and developing the schema for representing what is known in the world as knowledge in ways that computer can understand and store as information indexes. The objective of this paper is to review the concept theory which represents the schema for knowledge organization in Social Sciences. Consequently, this study needs to address the problem oriented schema for Social Science Information.

はじめに

概念理論とは厳めしい名前であるが、その目的はデジタルメディアの情報を明らかにするプログラム設計論のことである。これを別名、計算可能意味論 [齊藤 2006a] と呼ぶこともできる。名前のとおり意味の解釈をコンピュータ処理（アルゴリズム化）できないものかを研究する。その応用となるものが情報インデックスである。本稿の情報インデックスは社会科学に関する図書文献、音声と映像、さらにデジタル化されたマルチメディアなどの記録情報を分類する知識組織化に使用される。そして完成するものはデータベースや Semantic Web と呼ぶ社会科学に関わる情報システムである。情報の主題分析は、まず視点を定めて主題の抽出から始まり、主題を構成する概念構造をスキーマとして明らかにし、情報インデックスによってプログラムされる。スキーマという概念文法はアリストテレス・ペーコン流儀の古典的な分類も含まれ、またランガナータン・ビッカリー流儀のファセット (Facet) の考え方も同類といえる。これらは汎用のスキーマであり、主題の平均像を掴み取る視点として参考になるが、主題分析を深化する道具としては強力なものではない。研究者の視点に基づく社会科学情報に適した問題向きスキーマを設計する必要があるから、そのために考慮すべき概念理論について展望する。

1 主題分析

情報の組織化を主題分析というが、それは記録の中に刷り込まれた主題の姿をいかにして掴み取るかという技術であり、これまで様々な理論が提案されてきた。[斉藤 2006c]

記録の主題は内容から影響されるものであるが、内容そのもとと区別しなければならない。内容の多くは自然言語を用いて記述されるのに対し、主題は概念言語を用いて表される。両者の違いを簡単な事例を用いて明らかにしてみよう。ここに「斉藤孝 中央大学 情報学」と「今井悠真 東京大学 医学」の2つの記録があるとする。これらの記録は、そのままでは漢字の文字列(データ)にすぎなく、何を意味する記録であるかは不明である。もし誰かが「氏名 所属 専攻」という項目を考えて、記録に対応付けたとする。彼は「名簿」を表すものではないかという概念を頭に描いたからだ。その概念こそが主題であり、記録に意味を与えるものである。それは「名簿(氏名, 所属, 専攻)」という枠組みとして主題の構造を表す。また1つの情報化の視点となったものであることからスキーマと呼ぶ。このスキーマは、「名簿(氏名, 所属, 専攻)」という概念言語によって定義されたものである。[斉藤 2004]

このような主題分析を経て概念言語を用いて記録情報は処理される。この事例では「氏名(斉藤孝, 今井悠真), 所属(中央大学, 東京大学), 専攻(情報学, 医学)」という分類が可能になり、それぞれの文字データは意味を持つことになる。「氏名, 所属, 専攻」という分類項目(カテゴリ)は概念名となり、その上位概念である「名簿」を頂点とした階層構造となることからタクソミーと呼ぶ。また同時に集合論を用いて概念をクラスと呼び、その要素となるものをインスタンスと呼ぶこともできる。「氏名(斉藤孝, 今井悠真)」では、「氏名」というクラスには斉藤孝と今井悠真というインスタンス(実体)が属している。クラスとインスタンスの関係はメタ言語に準拠した場合、メタデータとデータとの関係にそのまま対応できる。まとめると、概念言語とはタクソミー(分類)という概念文法に裏付けられた概念語から構成されたものである。主題分析の視点をいかにして設定するか、そしてスキーマをいかにして把握するか、そのためには知識がなければならない。新聞記事を書く記者は、「5W1H」を常に頭に描くことを心がける。これとて「Who, What, Where, When, Why, How」という記事を取材するスキーマである。スキーマは概念の枠組みであるから、構造を持つことが特色である。その構造は階層構造や木構造で表すものもあれば、網構造で表すものもある。いずれにしても上位・下位、包含などの概念間の関係性を示す規則を持つ概念文法となる。[斉藤 2010c]

1.1 情報インデックス

普通のインデックスは、ある対象を指示するリンクの役割を持ち、索引語などから逆引きすることによって対象主題の文脈にアクセスする仕組みである。[斉藤 2010b] それは荷物とその

荷札との関係に似たものである。言葉では、名辞とそれが指示する概念との関係に相当し、ある概念に対して様々な名前を付与することもインデックスになるだろう。データベースやデジタル・ライブラリなどの情報インデックスは、対象主題を表す概念語彙（索引語のインスタンスとクラス）と索引辞書（統制主題記述語彙）とを関係付けることによって上位下位などの概念階層を明らかにする。そこには概念への深い思索がある。

ところで誰が情報インデックスを作るのか、それは一般にインデッカーの仕事とされるが、Web時代では誰もがインデッカーになれる。何故ならばフォクソノミー（Folksonomy）に見られるような自律的情報発信 [斉藤 2008] の時代であるからだ。インデッカーは昔のような情報に精通した専門職を表す名前ではなく、Webに接する誰もが担える知的な行為者といえるだろう。インデッカーは概念言語を操り情報の発信と検索を行う。情報インデックス処理の手順をまとめると list 1 になる。 [斉藤 2012]

[list 1] 情報インデックス処理の手順

- 手順 1 利用者とデジタルメディアが存在する。
- 手順 2 そこに利用者の暗黙知のスキーマ（概念文法）が加わる。
- 手順 3 インデッカーの概念設計により形式知のスキーマが整備される。
- 手順 4 インデッカーはタクソノミーに基づき情報インデックスを作成する。
- 手順 5 利用者はスキーマと情報インデックスに支援されて、知識地図（Kmap）を描く。
- 手順 6 情報インデックスはタクソノミーの支援によって情報共有される。
- 手順 7 情報インデックスは、陽のメタデータを用いて情報の明示化を促進する。

利用者は、主題のスキーマとそれを支援する情報インデックスの2つによってデジタルメディアに接する。主題のスキーマは陰のメタデータによって与えられた情報の窓である。このスキーマを通して利用者はデジタルメディアの基本的な概念を明らかにすることができる。例えば、それが「書籍」という主題構造を持つものであるとか、それが「新聞」という主題構造を持つものであるという基本的なりテラシーに支えられた暗黙知のスキーマである。日常生活でも知的な関心を持つ時、無意識に暗黙知のスキーマという概念文法が働く。その中ではリテラシーの基盤とされる書籍スキーマが重要である。書籍という仕組みには何らかの知識と体系を押し込むコンテナの機能が見られる。書籍の属性には、形式化という共通基盤（プロトコルとアルゴリズム）が重要な要素となる。つまり、人間が様々な知識、様々な情報、それをいかにして形のあるものにし、それを明示化するか、そして相手に理解してもらうか、そのためにできるだけ多くの人々に理解されるために共通の書籍スキーマを利用する。書籍スキーマは情報の窓でもあり、その窓からいかにして知識を得るのか、また得られるようにするのか、書籍

スキーマは、リテラシーの歴史に支えられた普遍的な書籍を理解する概念文法となった。これはまた陰のメタデータ [斉藤 2007b] といえるものである。利用者にとって陰のメタデータだけでは不十分である。多くの人々と情報を共有し、普遍的な知識を体系化するために概念言語が必要になる。それは陽のメタデータとして設計される。この2つの陰と陽のメタデータで支えられたスキーマを用いて利用者は、デジタルメディアと向かい合い、知的な相互作用を行う。そして利用者は、知識地図 (Kmap) を描き始める。[斉藤 2010c] それは知識の進化を表すもので、デジタルメディアを使えば使うほどに知識地図は成長していく。Kmap は利用者の知識の推移を示すことができ、その出発点となるものが原始スキーマである。そのスキーマは知識を形あるものにするために、その雛型 (モデル) を利用者にとり与えてくれる。[斉藤 2009b]

1.2 スキーマと情報インデックス

ここでは、社会情報学を用いてそのスキーマと情報インデックスについて説明する。社会情報学の主題分析は3つの視点から行われている。これは社会情報学の定義を表すものではなく、計量言語学的方法 [斉藤 2006b] から導かれたもので、実際の文献の中で使用された語彙調査に基づくものである。3つの視点は、概略的なものではあるが社会情報学のスキーマを明らかにする。[斉藤 2007c]

- (A) コミュニケーションの視点
- (B) アプリケーションの視点
- (C) ドキュメンテーションの視点

(A) は、社会調査、社会心理学、アンケート、インタビューなどにより要求を分析し、情報化する視点である。(B) は、図書館、情報センター、情報システム、社会学、法学、経済学、文学、医学などの学問的知識、ビジネス知識などの応用能力に関する視点である。(C) は、研究に関わる資料を作成する記録情報化の視点である。どのような学問であっても、そこには必ずそれを行う人間と、その人間のおかれた場面が存在する。その個々の人間と場面により視点が決定される。その上で対象となる学問が研究される。学問という対象を分析する際、これらの要素を加味しなければ全体像は見えてこないはずである。したがって、人間を「Agent」、視点を「View」とし、対象つまり「Object」を研究するというスキーマを見出すことができる。

[list 2] 社会情報学の主題とスキーマ

- (1) 社会情報学 (理論, 歴史, コミュニケーション, 情報メディア, 社会調査, 統計分析, 情報処理)
- (2) 社会情報学 (Agent (研究者, 学生, 市民), View (研究, 卒論, 女性問題), Methodology

(理論, 社会調査, 統計処理, 情報処理), Object (社会, 情報, 経済, 政治, 文化, 歴史, コミュニケーション, 情報メディア))

list 2 (1) は, 1つのレベルの階層だけの単純な分類で社会情報学の主題を表しているだけで, スキーマが明らかでないから, その構造は曖昧である. そこで list 2 (2) のように「Agent・View・Methodology・Object」という概念言語によるスキーマを与えてみる. すると主体 (Agent) がいて, ある場面 (View) において, ある方法論 (Methodology) を用いて, ある対象 (Object) を処理するという主題の構造が鮮明になる. このようなスキーマの定義によって list 3 と list 4 のように情報インデックスとして処理される. [齊藤 2010d]

[list 3] スキーマのプログラムの例

```
<class> 社会情報学 </class>
<facet id= "Agent"> 主体 </facet>
<facet id= "View"> 場面 </facet>
<facet id= "Methodology"> 方法論 </facet>
<facet id= "Object"> 対象 </facet>
<topic id= "kenkyusa" facet id= "Agent"><name> 研究者 </name></topic>
<topic id= "gakusei" facet id= "Agent"><name> 学生 </name></topic>
<topic id= "shimin" facet id= "Agent"><name> 市民 </name></topic>
<topic id= "kenkyu" facet id= "View"><name> 研究 </name></topic>
<topic id= "sotsuron" facet id= "View"><name> 卒論 </name></topic>
<topic id= "jyoseimondai" facet id= "View"><name> 女性問題 </name></topic>
<topic id= "riron" facet id= "Methodology"><name> 理論 </name></topic>
<topic id= "syakaichosa" facet id= "Methodology"><name> 社会調査 </name></topic>
<topic id= "tokeisyori" facet id= "Methodology"><name> 統計処理 </name></topic>
<topic id= "johosyori" facet id= "Methodology"><name> 情報処理 </name></topic>
```

[list 4] 情報インデキシングの例

```
<page url= "http://www.saitolab.chuo-u.ac.jp/SocioInfoSys/">
<title> 社会調査に関わる女性問題のポータルサイト </title>
<description> そのポータルサイトは, 情報主題 (サブジェクト) として「社会調査に関わる女性問題」を扱う. その情報システムは, Web リソース (記録) から定期的にその主題に関わるリソースを収集して, 記録を情報化する. そして, 研究者に対して情報処理のためのイ
```

```
ンタフェースを提供する. </description>
<occurrence topic id= “社会調査” />
<occurrence topic id= “女性問題” />
<occurrence topic id= “研究者” />
<occurrence topic id= “情報処理” />
</page>
```

list 4 は、「社会調査に関わる女性問題のポータルサイト」と題する記録を情報インデックスしたものである。それは、社会調査、女性問題、研究者、情報処理の4つの語彙を抽出し、それぞれの topic id を <occurrence> というタグによって付与されている。なお、ここではタグや RDF [Blank 2004] などのメタデータ言語 [齊藤 2009c] の説明は省いている。

1.3 オントロジのスキーマ

オントロジでは概念間の関係性 [齊藤 2005] に注目するが、その場合のスキーマの把握について説明する。次のような記録があったとする。

「農業は、土地の性質や気候条件などの、自然の力に影響されやすい産業です。したがって、人間の力や機械の力で品物を生産している工業とはちがいで、自然の災害を受けやすいのです。特に、日本は、国土の位置が、北の寒い地方から南のあたたかい地方にまで、弓のような形に細長くのびているので、気候の影響を受けやすく、農業災害がおこりやすい国です。農業災害のおもなものとしては、干害（ひでりの害）や冷害をはじめ、霜害（しもの害）・風水害・病虫害・ひょうの害などがあります。」(菊地家達筆、『社会科事典3——農林業・水産業——』。国土社。1994。36 ページ)

この主題は、「日本の農業災害」であり、それには「干害、冷害、霜害、風水害、病虫害、ひょうの害」などの種類があるという。まず日本を「Place」として掴み取る。そこにおけるクラスを「Disaster」とし、そのインスタンスを「干害、冷害、霜害、風水害、病虫害、ひょうの害」とすれば、「Disaster = 農業災害」というスキーマが明らかになる。さらにオントロジの視点から「Place」と「Disaster」の2つの概念の間には「起こる (cause)」という関係性を導き出す。定義されたスキーマは list 5 (1) である。そのまま概念言語として、またメタデータ言語として list 5 (2) の情報インデックスに利用できる。

[list 5] オントロジのスキーマ

(1) 災害のスキーマ

subject (cause (Disaster, Place))

(2) 情報インデックスの例

<subject>

<cause>

<Disaster> 干害 </Disaster>

<Disaster> 冷害 </Disaster>

<Disaster> 霜害 </Disaster>

<Disaster> 風水害 </Disaster>

<Disaster> 病虫害 </Disaster>

<Disaster> ひょうの害 </Disaster>

<Place> 日本 </Place>

</cause>

</subject>

このようなオントロジによるスキーマをもう1つ見てみる。実際の新聞記事を対象にしたものである。[齊藤 2006c]

「横須賀市の市立神明小で26日、新聞記事を使った社会科の授業があった。同市内の公立校教諭らでつくる市新聞教育研究会が、新聞に児童を親しませようという授業研究としておこなったもので、同市では初めての試み。モデルとなった4年2組の教室では、36人の児童たちが活発に意見を出し合い、45分の授業時間を超えるほど。参加した約40人の先生たちも、ほっとした表情だった。同市教委に科目別などに置かれた研究会の1つで、これまで「新聞を教育に使うことで、子供をどう育てるか」をテーマに、討議を重ねてきた。この日授業を受け持った田野口則子教諭はまず、今年1、2月の各紙から、日本各地の生活の様子が見える記事を児童に選ばせた。続いて、それらの特徴を挙げさせる。子供たちは「北海道の流水」や「沖縄の満開に近いサクラ」といった記事から、活発に意見を述べあった。」(『朝日新聞』1992年02月27日 朝刊 神奈川)

この新聞記事では、田野口則子教諭が36人の児童に社会科を教えているから、“誰か”に関する概念は教師 (Teacher) とする。そして“誰かに”に関する概念は生徒 (Student) とする。

さらに“何かを”に関する概念は科目 (Curriculum) とする。この主題は「小学校の授業」というもので、「Teacher, Student, Curriculum」という3つの概念 (クラス) によって成り立つ。次の処理は概念間の関係性の定義である。関係性として、教師は生徒に授業を行っており、ここでは「coach」という関係子を用いて、「Teacher」と「Student」とを関係付ける。「teach」としないのは、「Teacher」と混同しないためである。教師は科目を担当しているということである。ここでは「charge」という関係子を定義して、「Teacher」と「Curriculum」とを関係付ける。教師は道具を用いて授業を進めている。ここで「use」という関係子を定義して、「Teacher」と「Tool」とを関係付ける。道具という概念は、「isa」を用いて「Tool」と「Newspaper」とを関係付ける。定義されたスキーマは list 6 である。

[list 6] 新聞記事のスキーマ

use (Teacher, Tool), charge (Teacher, Curriculum), coach (Teacher, Student)

isa (Newspaper, Tool)

2 概念理論

スキーマの設計には概念理論が必要である。1879年ドイツの論理学者フレーゲは、『概念表記法』と題する書物を出版したが、それはアリストテレス以来の二千年にわたる形式論理学の伝統を破るものとされた。フレーゲは、論理とは真理に関わるもので、精神が生物的な影響や新しい環境がどのように変化しても真理の法則は進化することなく永久に不変のままであると主張したのである。たしかにフレーゲの概念表記法は、精神を言葉の支配から解放するためとして後世になり高く評価された。それまでのアリストテレス流儀の古典論理学では、「三段論法」に見られるように概念の定義には自然言語を使っていたが、それに代わり概念表記法 (Concept Script) という独特の記号を用いて「ある」、「任意の」など量化推論の形式化を厳密に表すことができた。フレーゲの難解な概念言語は使われることはなかったが、概念理論として人々を魅了させるものとなった。

情報学の概念理論を包括的に展望したのは、オクスフォード大学の Floridi [Floridi 2011] である。彼はそれを PI (Philosophy of Information) と呼び、原点の哲学から先端の IT に至る概念理論について系統的に考察した。情報学の立場から概念理論を丁寧にたどる研究者はデンマークの Hjørland である。彼の論文 [Hjørland 2002] [Hjørland 2007] は汲み取らなければいけない論点を指摘する。

2.1 記号論

概念を科学的研究対象にしたのは記号論である。記号論はこのように展開される。人間は、