

知識メディアシステム CM-2 とそのユーザインタフェース

前田 晴美・西田 豊明

(奈良先端科学技術大学院大学 情報科学研究科)

A Knowledge Media System and its User-Interface

Harumi MAEDA, Toyoaki NISHIDA

Graduate School of Information Science,

Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara, 630-01 Japan

Abstract: In this paper, we present a knowledge media system *CM-2* (Contextual Media system version 2) which provides users with a means of accumulating, sharing, and exploring promiscuous information gathered from vast information sources. Our approach is based on the idea of *knowledge medium* using *associative representation*. We propose that the method of describing any relationship in the universe of discourse using associative representation is plain and therefore powerful to gather information from natural information sources. We show some examples of how *CM-2* can support user activities in research work.

keyword: Associative Representation; Knowledge Media; Knowledge Media System; *CM-2*; User-Interface; Information Retrieval.

e-mail: harumi-m@is.aist-nara.ac.jp

1 はじめに

研究活動にはさまざまな知識や情報が必要である。例えば、自分が過去に作成した論文やOHP、アイディアメモなどを中心に、研究動向を調査するための書籍・雑誌、広く知識を収集するための辞典や辞書、オンラインデータベースやWWW(World Wide Web)など、種類も形態も異なる雑多な情報である。文書化・電子化されていない頭の中にあるアイディアの断片や常識なども非常に重要である。

できるだけ手間をかけずに知識や情報の収集、整理を行い、創造的な思考活動に集中したいというのが研究者共通の要求である。そのためには、世の中に存在するさまざまな種類の知識情報を簡単に収集、整理でき、それらを利用して論文作成などの知的作業に応用できるようなシステムの実現が望まれる。

しかし、現在のコンピュータ支援環境には以下のような問題点がある。(1) 内容、形式や構造がまちまちな情報を整理するには、RDBやOODBなどのデータベース構築や、一階述語論理などによる知識ベース構築は煩わしい。(2) 情報構造が個々の情報によって違うWWWでは、自分のほしい情報を探すことはたいへんであり、収集した情報を使いやすいように選別、加工することは人手で行わなければならず面倒である。

なぜ既存のデータベース・知識ベース構築が煩わしいか考えると、関係モデルや一階述語論理などの知識情報表現がコンピュータ向きで強力なデータ管理や推論などを実現できる反面、人間にとって理解しにくく、利用者が手軽に知識情報を入力しにくいことがあげられる。またWWWに代表されるマルチメディア情報は人間には理解しやすいが、情報の持つ意味を扱うことができないため、コンピュータ

が人間のかわりに情報の加工をすることが難しい。

そこで我々は、知識情報を RDB や OODB よりも簡単に整理・収集して、WWW よりも容易に利用するために、人間向きメディアとコンピュータ向きメディアの中間表現として、連想表現を基本構造とする知識メディア [1] を提案する [2,3,4]。

連想表現は概念間を連想関係で結ぶ知識表現の方法である。連想表現を基本構造とする知識メディアは、従来の知識情報表現と比較するとはるかに単純で人間向きであり、マルチメディア情報と比べるとコンピュータにも扱いやすい。すべての概念関係を連想関係にまとめる方法では、既存技術で利用可能な機能の全ては実現できないかもしれないが、人間の負荷をできるだけ減らし、短時間で情報の収集、整理を行うためのトレードオフとして有効であると我々は考える。

我々は本アプローチに基づき知識メディアシステム CM-2 (Contextual Media System version 2) を試作している。CM-2 には従来のシステムにはない新しいしくみが必要であり、以下の手法を検討し、部分的に開発を行った。

- 自然界に存在する内容も構造も違うさまざまな情報を取り込むための情報キャプチャ機構
- 情報を特定の視点から収集・利用するための知的統合機構
- 連想に基づく情報ベースから必要な情報を検索するための連想検索機構

以下では、2章で CM-2 の概要とアーキテクチャについて紹介し、3、4、5章で CM-2 の3つの手法について説明する。6章では、実際の研究活動への CM-2 の適用例を紹介する。

2 CM-2

2.1 CM-2 の概要

CM-2 は情報ベース (information base) の集合から構成される。情報ベースのユーザインタフェースの中心となるのはカード型の作業空間 (workspace) とエージェント (agent) である。利用者は、日常ノート上で自由にメモを書いたり考えをまとめたりする感覚で、作業空間上で思考活動や情報収集などのさまざまな情報活動を行う。作業空間には自律的なソフトウェアモジュールであるエージェントが住んでおりさまざまな情報サービスを提供する。利用者やエージェントはネットワークに接続された情報ベースや他の情報源を利用することができる。

Fig1に CM-2 のイメージ図を示す。

2.2 連想表現を基本構造とする知識メディア

連想表現を基本構造とする知識メディアは、概念 (concept) 間の連想 (association) による連結を基本構造とする弱構造

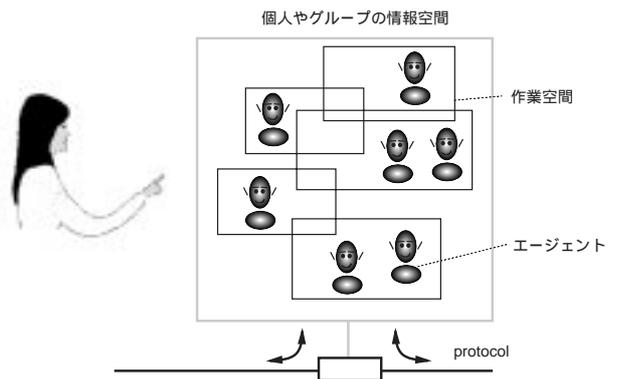


Fig 1: CM-2 のイメージ図

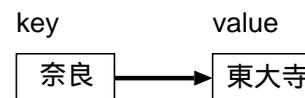


Fig 2: CM-2 における連想表現

のメディアである。知識メディアの最小単位をユニットと呼ぶ。以下に主要なユニットの種類を示す。

- 概念 (concept) : 概念ユニット
- 外部参照データ (external-data) : 外部情報を参照するユニット
- 連想 (association) : 概念または外部参照データ間を連想関係で連結するユニット

1つ以上の概念または外部参照データから1つ以上の概念または外部参照データが連想される。連想元を key、連想先を value と呼ぶ。本稿では CM-2 における連想表現を Fig2 のように記述する。この例では key 「奈良」から value 「東大寺」が連想されている。

CM-2 では、「奈良県」「奈良市」のような全体・部分関係、「寺」「東大寺」のようなクラス・インスタンス関係などの概念間のさまざまな関係を、連想表現を用いて記述する。そうすることにより人間が関係の分類に悩むことなく情報ベースを作成することができる。

2.3 システム構成

CM-2 は、内部知識ベースと外部データベースから構成される情報ベースを中心とした、マルチメディア/ネットワーク対応のサーバ・クライアント型システムである。利用者は作業空間を通してネットワークに接続された個人やグループの情報ベースや、他の情報ベースを利用できる。Fig3にシステム構成を示す。

CM-2 は Common Lisp で記述され、GUI は Gcl/Tk を使用している。

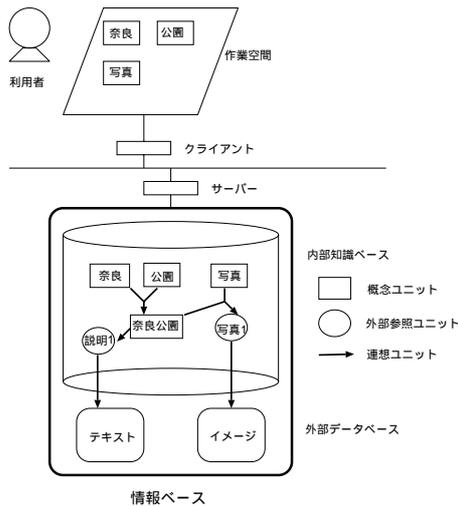


Fig 3: システム構成

3 情報キャプチャ機構

利用したい情報は、書籍や個人的な記憶、WWW などさまざまなところに存在する。自然界に存在する内容も構造も違う情報源から情報を取り込み、情報ベースを生成するための機構である。

自然情報源からの情報取り込みには膨大なヒューリスティックが必要であることが予測される。我々はWWW に焦点をあて、情報キャプチャの手法を検討する。

3.1 WWW からの知識情報取り込み

WWW の記述言語である HTML はかなり柔軟な言語であり、作者に厳格な情報構造化を要求しない。例えば <h1> などの見出し表現や や <ui> などの段落表現を使って文書を構造化してもよいし、しなくてもよい。また、見出し表現は、見出しとしてではなく、単にブラウザ表示時の文字の大きさを変えたいために使用されることもある。

以上のことから、HTML 文書のどの部分に利用者にとって有用な知識情報が書かれているかは自明ではない。我々は、いくつかのヒューリスティックを用いて WWW からの情報キャプチャの実験を行った。

3.2 実験

奈良先端科学技術大学院大学のホームページ¹ から出発して、アンカー表現に基づいた 1 段階の幅優先探索を行い、11 のヒューリスティックに基づき HTML 文書を取得・解析し、HTML 文書毎に小さな情報ベースを作成した。本実験でのヒューリスティックは見出し表現と段落表現に着目している。生成されたユニットに対して、適否を主観により評価した。

実験の結果、生成された全ユニット 342 個の内、適当なものは 224 個 (約 65%) であり、間違いではないが問題があ

るユニットも含めると 295 個 (約 86%) となり、概ね良好であると評価する。

不適当なユニットの 87% は、奈良先端科学技術大学院大学に直接的に関係のない情報 (例: WWW の使い方など) であった。また、「Click here for the home page in English.」のように、WWW に特徴的なハイパーリンク表現を概念ユニットとして抽出した例が 2 つあった。

間違いではないが問題がある概念ユニットの 87% が「組織 (組織図・役職員)」のような複合語であり、value として検索するにはよいが、key とするには分解したほうがよいものであった。また、<dd> 部からユニットを抽出する際、内容が概念というよりも、<dt> 部から抽出した概念の説明文であり、外部参照ユニットとしたほうが良いものが 11 例あった。

今回は奈良先端科学技術大学院大学のホームページという、比較的構造的に設計された HTML 文書からの取り込み実験を行ったため、結果は良好であった。今後は違うサイトで HTML 文書の数をややして実験を行い、アルゴリズムの調整を行う。

4 知的統合機構

個人やグループの作成した情報ベースや他の情報ベースのようなさまざまな情報ベースの情報を、特定の視点から収集して利用するための知的統合機構である。

4.1 同名統合アルゴリズム

現在のアルゴリズムは、概念ユニットの同名統合に基づいている。

1. 同名の概念ユニットがある場合同じ概念として扱う。
2. 連想ユニットを新しく再構成する。この際、ヒューリスティックを用いて元の情報ベースの視点に対応しない連想ユニットは削除する。

Fig4に例を示す。個人やグループの視点は作業空間として実現される。視点を表す作業空間 A に、「東大寺」から「お水取り」が連想され、「お水取り」から「春」が連想されることが記述されている。統合したい情報ベース B には「お水取り」から「東大寺」と「二月堂」が連想されることが記述されている。視点 A に基づいて情報ベース B を統合すると、連想ユニットの動的なくみかえおこり、「東大寺」から「お水取り」、「お水取り」から「春」と「二月堂」が連想される記述を持つ新しい情報ベース C が生成される。

4.2 実験

3.2の実験で作成した情報ベースの内、不適当なユニットを削除した 14 個の情報ベースを Fig5に示される視点から統合する実験を行った。

実験の結果、12 の概念ユニットが同じ概念と扱われ、連想ユニットのくみかえが行われた。主観的に評価したとこ

¹URL <http://www.aist-nara.ac.jp/>

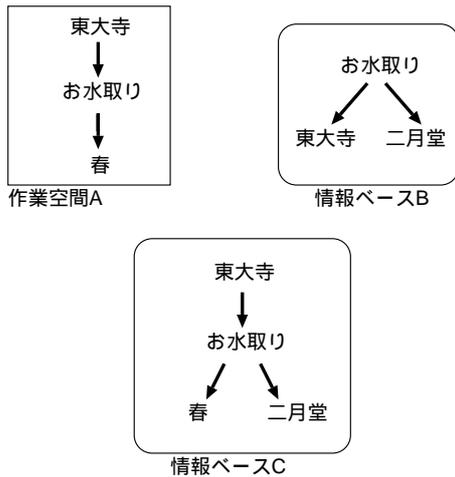


Fig 4: 視点に基づいた統合

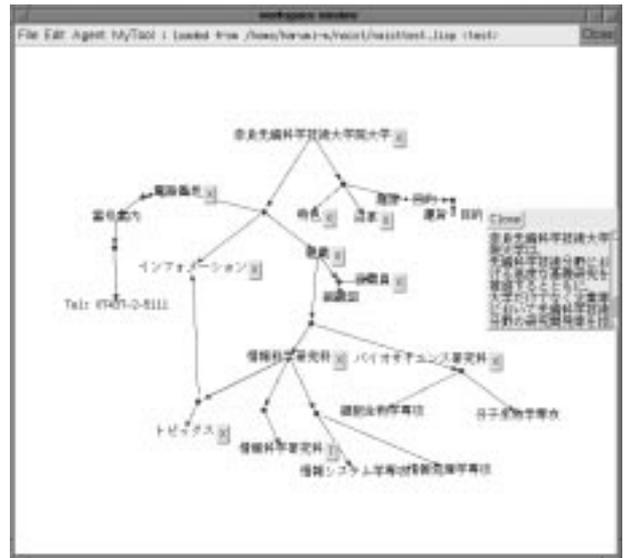


Fig 6: 統合後の視点



Fig 5: 視点

る統合された情報ベースの内容は良好である。統合後の情報ベースが Fig5の視点から構造化されている (Fig6参照)。

現在の知的統合機構では、概念の名前が完全一致しないと同じ概念ユニットとは見えないため、人によって概念に対する名前付けが違う場合や、概念が別名を持つ場合などに問題が生じる。

4.3 辞書機構

上記の問題点を解決するために、辞書機構を設計した。

辞書を表す作業空間に連想表現を用いて、「NAIST」「奈良先端科学技術大学院大学」のような同義語関係を記述する。視点に「NAIST」、情報ベースに「奈良先端科学技術大学院大学」と記述されている場合、辞書を参照してこれらの2つの概念を同じものとして扱うことができる。

5 連想検索機構

連想に基づく情報ベースから必要な情報を検索するための連想検索機構である。

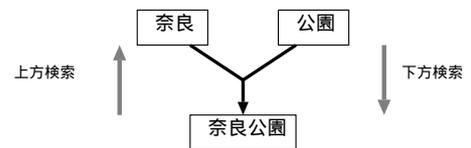


Fig 7: 近傍検索

5.1 近傍検索機構

概念の近傍を検索する。上方検索と下方検索の2種類がある。

- 上方検索: value から key を検索する
- 下方検索: key から value を検索する

例えば Fig7のような連想がある場合、上方検索では「奈良公園」から「奈良」「公園」を得、下方検索では「奈良」と「公園」から「奈良公園」を得ることができる。

5.2 活性境界探索型連想検索機構

概念の近傍を1つずつたどらなくても、概念を選択するとそこから連想されるものを情報ベースからダイナミックに検索する方法が活性境界探索型連想検索機構である。これは、Quillian[5]の活性伝播 (spreading activation) に基づいている。複数の概念を選択すると、その周囲に活性化の輪が広がり、輪が接触したところの概念を検索できる。

Fig8の例は、「奈良」と「寺」と「大仏」を与えると、概念の距離1 (連想によって直接接続されている) のところで活性化の輪が接触し、「東大寺」が解として得られることを示している。

本機構により、利用者が情報ベースの構造を知らなくても、「奈良で大仏がある寺は？」というような検索を行うことができる。

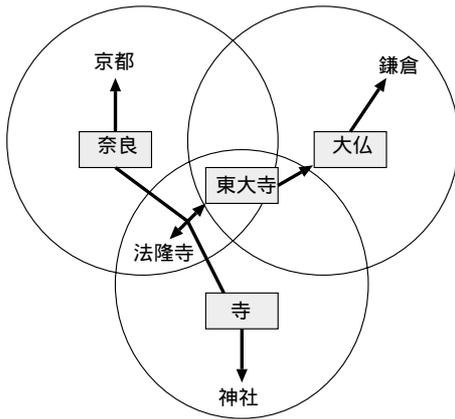


Fig 8: 活性境界探索型連想検索

5.3 実験

研究室で手作業で構築した奈良観光情報ベース²に対して、活性境界探索型連想検索を実験した。情報ベースには全部で約 2000 個のユニットを含む。質問は全部で 30 問である。ただし、元々の情報ベースの連想ユニットからすぐに答えが見つけれられるような質問や情報ベースに解の存在しないような質問は削除した。探索は最初は距離 1 で行い、正答が得られない場合のみ距離 2 探索を行った。

1 つの質問に対する評価は以下の 3 種類とした。

- すべての解が正解（事実と合う）である場合：正答（例：「奈良」と「寺」「東大寺」と「法隆寺」）
- 解なしあるいは解に正解が含まれない場合：誤答
- 解に正解及び事実と合わない解が含まれる場合：その他（例：「奈良」と「寺」「東大寺」と「神社」）

距離 1 探索での正答は 16/30(約 53%) である。距離 1 と距離 2 の総合評価では正答は 17/30(約 57%) である。正答にその他を含めると、正答率約 80% となる。

本結果は探索の輪をひろげても正答があまりふえないことを示している。ただし、少々関係ない答が出てきてもよいがとにかく正解がほしいような時に距離 2 探索は有効である。

また、実験を通じて、概念の数を増やせば正解が出やすくなるわけではないことがわかった。これは現在のアルゴリズムが概念の近傍の積集合をとるに基づいているため、和集合をとったり、活性化の輪の距離を概念毎に適切に調節するなどの工夫が必要である。

6 CM-2 による研究活動支援

我々の研究グループでは、実際に CM-2 を研究活動に利用している。現在はディスクや WWW など電子化された情

² 試験的に WWW にて公開中。

URL http://flute.aist-nara.ac.jp/KC/server_demo.html

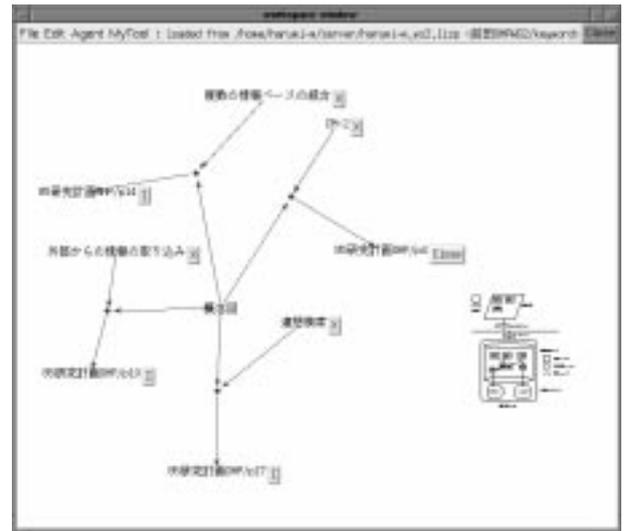


Fig 9: 情報整理支援 1

報を中心に情報を整理している。以下に例を示す。

6.1 情報整理支援

研究室のディスクの中には、思いつきを記したメモや、過去に作成したプログラムや論文・OHP など、さまざまな情報が存在する。これらの情報は雑多であるが、研究活動にとって非常に大切である。

CM-2 では情報整理を簡単に行うための枠組を提供する。

最もシンプルな方法は、整理したいドキュメントファイルにキーワードをつけ、キーワードとドキュメントファイルや、キーワード間を連想関係で結ぶことである。

Fig9の例は、本稿で使いそうな図表を、過去著者らが作成した OHP の中から探し出している様子を示す。キーワードを画面から入力または選択すると連想記憶をたどるように、関連する情報が次々と検索される。この場合は、「CM-2」と「概念図」から連想される外部参照ユニット「研究計画 OHP/p4」の横のボタンをマウスで選択すると OHP の縮小画像が表示されている。

さらに便利な方法として、ディスクの中の情報をさまざまな観点から使いやすく整理するエージェントの開発を行っている。

Fig10では、エージェントが利用者からの指示により電子メールを分類している様子を示す。エージェントを活性化することにより整理が行われる。

現在のエージェントは利用者とインタラクティブに対話をしているが、将来的には利用者の好みや使用頻度と傾向を学習するなどして自律的に情報の整理が行えるエージェントを開発する予定である。

6.2 情報分析支援

情報を分析してわかりやすく提示することは人間の理解を促進する。CM-2 ではグラフや図表などの人間にとって

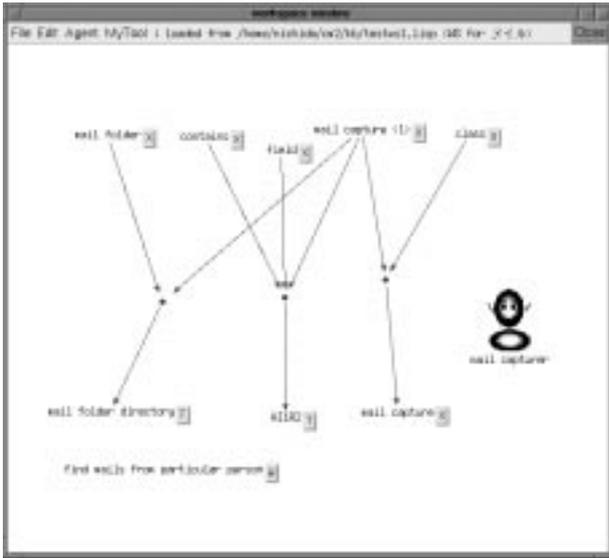


Fig 10: 情報整理支援 2

わかりやすい表現を用いて情報内容の分析を行うエージェントを開発している。

Fig11は、CM-2 プログラム構造図の一部を示している。Common Lisp プログラムを解読するエージェントが、ディレクトリ/ファイルが与えられるとその中に含まれる関数の呼応関係を抽出して提示する。

このほか、論文やフォーマットされたドキュメントを読んで要約や説明をするエージェントを開発する予定である。

7 おわりに

本稿では、研究活動に必要なさまざまな自然情報源の情報を有効に活用するために、人間向きメディアとコンピュータ向きメディアの中間表現として、連想表現を基本構造と

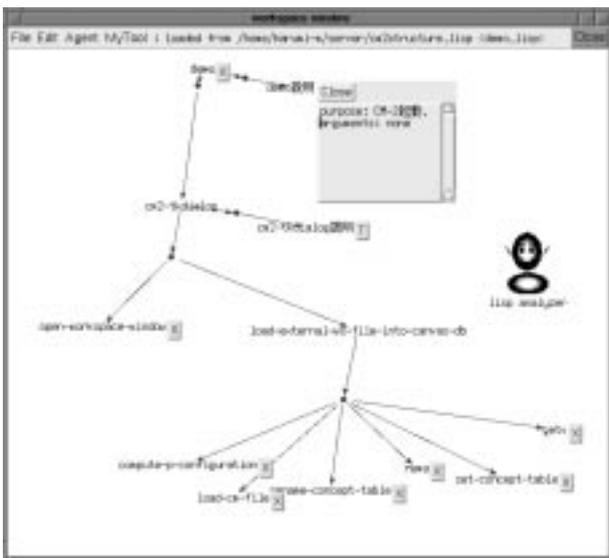


Fig 11: 情報分析支援

する知識メディアを提案した。本アプローチに基づいて知識メディアシステム CM-2 を試作し、実際に研究活動に利用した。本研究で用いたアプローチは研究活動だけではなく、個人やグループベースでの知的生産活動においても有用であると考えます。今後は、CM-2 の開発を継続して進めながら、他の事例への適用研究も行う予定である。

参考文献

- [1]M.Stefik. The next Knowledge Medium. AI Magazine, Vol.7, No.1, pp.34-46, 1986
- [2]T.Nishida and H.Takeda. Towards the Knowledgeable Community. In Proceedings International Conference on Building and Sharing of Very-Large Scale Knowledge Bases '93 (KB&KS '93), Ohmsha Ltd. pp.157-166, 1993.
- [3]T.Nishida, K.Koujitani and H.Takeda. A Plain Indexing Method for Organizing Conceptually Promiscuous Data. AAAI Fall Symposium(予定), 1995.
- [4] 梶谷和人、前田晴美、西田豊明. 文脈メディアによる情報の提示と構造化. HIS'95(予定), 1995.
- [5]M.R.Quillian. Semantic Memory. in M.Minsky (Ed.), Semantic Information Processing, MIT Press, 1968.

著者略歴

前田 晴美

1986年 京都大学文学部哲学科（心理学専攻）卒業、同年富士通株式会社入社。
1992-1993年 英国マンチェスター科学技術大学計算機学科修士課程 (MSc. in Computation)。
1995年 奈良先端科学技術大学院大学情報科学研究科博士後期課程入学、現在に至る。

西田 豊明

1977年 京都大学工学部情報工学科卒業。
1979年 同大学院修士課程修了。
1980年 同大学博士課程退学、同年より京都大学工学部助手。
1988年 助教授。
1993年 奈良先端科学技術大学院大学情報科学研究科教授、現在に至る。
1984年 1年間 Yale 大学客員研究員。
1995年 科学技術庁金属材料研究所技術研究所客員研究官。京都大学工学博士。知識の共有と再利用、知識メディア、定性推論の研究に従事。
著書：自然言語処理入門（オーム社）、定性推論の諸相（朝倉書店）等。