ＫＰ化による論文内容の効果的提示方法とその応用

松尾 利行・武田 英明・西田 豊明
（奈良先端科学技術大学院大学 情報科学研究科）

An Application of Knowledge Piece to Represent Contents of Papers

Toshiyuki MATSUO, Hideaki TAKEDA, Toyoaki NISHIDA
Graduate School of Information Science,
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-01 Japan

In this paper, we apply Knowledge Medium System(KMS) to semi-automated summarizing of technical papers in metallurgy. Knowledge Piece(KP) which we propose is a very powerful and robust method, because it is a knowledge chunk which has an appropriate grain size to textual process. KP consists of several sentences and linkage attributes toward other KPs. We develop METIS(METallurgy papers Intelligent Surveyor) system which acquires KPs from technical papers in metallurgy, summarizes each paper and survey plural papers from KPs. We estimate METIS performance using technical papers with experts.

keyword: knowledge medium;knowledge representation;
e-mail:tosiy-m@is.aist-nara.ac.jp

1　はじめに

近年、論文情報が電子化されるに伴い、論文誌をネットワーク利用して提供したりあるいは電子媒体で電子的に出版しようとする動きが顕著になっている [3]。そのような環境の整備に伴い、電子的データベースを使った検索機能の導入など、これまで紙ベースの出版では得られなかった種々のサービスが提供可能になる。

そこで、電子化された論文から論文情報ベースを構築し、論文内容を効果的に提示することを本研究の目的とする。

最初に、従来の学術情報データベースの問題点を明確にし、専門家あるいは研究者が、これらの情報ベースに何を求めているか (要求仕様) を確定した。次に、電子化された論文情報から論文情報ベースを構築するために、金属材料論文を対象に論文の内部構造を分析した。その結果、この種の論文がかなり構造化されており、専門用語が出現する頻度が高いため、ヒューリスティックなテキスト処理により、論文情報ベースを自動的に構築できる見込みが得られた。

論文内容を効果的に提示する論文情報ベースを開発することで、従来の学術情報データベースに対する要求仕様のいくつかを解決できる。本研究では、論文内容の効果的提示方法を検討した後、金属材料論文情報ベースを自動的に構築し、論文内容を提示するシステム (METIS:METallurgy papers Intelligent Surveyor) を開発した。METIS は、マークアップ言語（ＨＴＭＬ）で書かれている論文を入力とし、論文の要約および論文のサーベイを出力することができる。

論文の要約としては、(1) 論文内容の構造化による要約 (Fig1,Fig2参照)、(2) 論文中の全図表のリンク結合による論文内容のマップ表現 (Fig3参照)、(3) 因果関係情報 (Fig4参照)、(4) 因果関係ネットワーク (Fig8参照)、(5) 金属特性・要因間の因果関係の一覧表形式による表現 (Table1参照)、

Effect of processing conditions and heat treatments on mechanical properties of single–crystal superalloy CMSX–2

SUMMARY_PATTERN1

1. super alloy
    1. MXON
    2. CMSX–2
    3. MAR–M200
    4. TMO
    5. NASAIR 100
2. experimental result
    1. creep property
        1. creep life
        2. creep behaviour
        3. creep primary
        4. creep secondary
        5. creep strength
        6. creep rupture life
        7. creep rate
    2. fatigue property
        1. fatigue limit
        2. fatigue strength
        3. fatigue low cycle
        4. fatigue high cycle
        5. fatigue failure
        6. fatigue crack
    3. corrosion property
    4. other property
3. discussion
    1. process
        1. hip
        2. prcoess
        3. temperature
        4. casting
        5. solidification
        6. heat treatment
        7. composition
        8. addtion

Fig 1: 論文の要約 1

Effect of processing conditions and heat treatments on mechanical properties of single–crystal superalloy CMSX–2

SUMMARY_PATTERN2

1. super alloy
    1. MXON
    2. CMSX–2
    3. MAR–M200
    4. TMO
    5. NASAIR 100
2. creep
    1. creep property
        1. creep life
        2. creep behaviour
        3. creep primary
        4. creep secondary
        5. creep strength
        6. creep rupture life
        7. creep rate
    2. discussion
        1. process
            1. solidification
            2. heat treatment
        2. structure
            1. single crystal
            2. misfit
            3. precipitate
            4. pore
            5. dislocation
            6. homogeneity
3. fatigue
    1. fatigue property
        1. fatigue limit
        2. fatigue strength
        3. fatigue low cycle
        4. fatigue high cycle
        5. fatigue failure
        6. fatigue crack

Fig 2: 論文の要約 2

(6) 因果関係の内容一覧 (Table3参照)、(7) 図表情報のテーブル表現がある。ただし、(2)(4) のみは現在まだ完全自動化はしていない。

また、複数論文のサーベイとして、(8) 類似度による論文の空間配置 (Fig6参照)、(9) 類似度の説明機能 (Table4参照)、(10) 3 次元表示による因果関係の分布 (Fig7参照) があり、これらはいずれも自動化されている。

本研究では、論文情報ベースを構築するために、ＫＰ (KnowledgePiece: 知識片) という概念を導入した。ＫＰとは、弱い構造をもった知識メディア (Stefik[2]) 表現の１種で、カードサイズ程度の情報量を持つ知識のチャンクであり、中間知識表現として適度な粒度である。

プロトタイプを評価するために、１０件の論文で筆者がテストした後、１５件について知識ベースを無修正のままブラインドテストを実施し、金属材料の専門家から下記の評価を受けた。(1) 論文内容のマップ表現は、理解を深めるのに効果が大きいが、表示上の工夫が必要である、(2) 不適切な使用をしている専門用語がある、など多少の問題点はあるが、論文要約と論文サーベイは、文献の検索や理解を助けるのに効果が大きい。

2 背景

学術分野で、多様なデータベースが構築されてきた。従来型システムの問題点を明確にし、学術情報データベースに対する要求仕様を明確にする。

2.1 学術情報データベース構築に関する問題点

学術情報データベースを構築するためには、一次情報データと二次情報データの 2 つが必要である。一次情報データはオリジナルな情報であり、二次情報データはそのような情報への案内となる情報である。二次情報は、作成時の索引付けや抄録作成といった人手による作業があるため、登

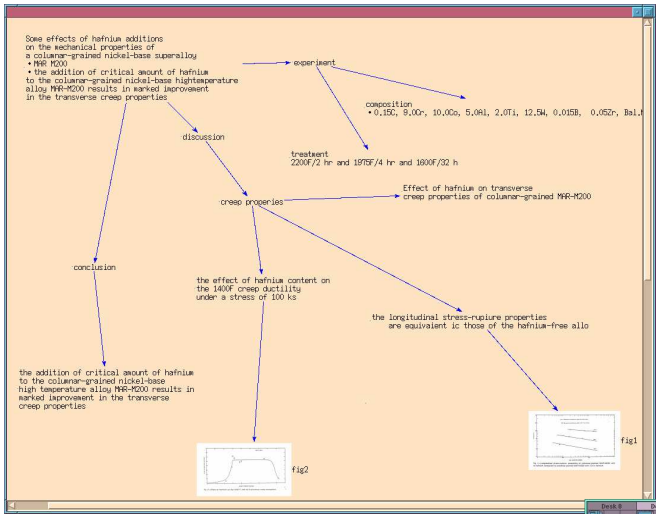
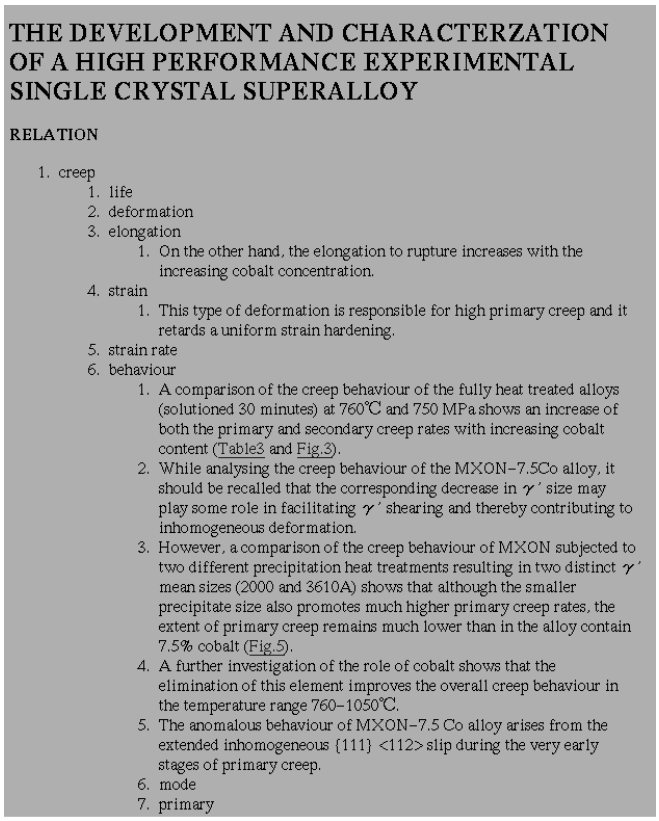


Fig 3: 論文中の全図表のリンク結合による論文内容のマップ表現



THE DEVELOPMENT AND CHARACTERZATION OF A HIGH PERFORMANCE EXPERIMENTAL SINGLE CRYSTAL SUPERALLOY

RELATION

1. creep
    1. life
    2. deformation
    3. elongation
        1. On the other hand, the elongation to rupture increases with the increasing cobalt concentration.
    4. strain
        1. This type of deformation is responsible for high primary creep and it retards a uniform strain hardening.
    5. strain rate
    6. behaviour
        1. A comparison of the creep behaviour of the fully heat treated alloys (solutioned 30 minutes) at 760℃ and 750 MPa shows an increase of both the primary and secondary creep rates with increasing cobalt content (Table3 and Fig.3).
        2. While analysing the creep behaviour of the MXON−7.5Co alloy, it should be recalled that the corresponding decrease in γ' size may play some role in facilitating γ' shearing and thereby contributing to inhomogeneous deformation.
        3. However, a comparison of the creep behaviour of MXON subjected to two different precipitation heat treatments resulting in two distinct γ' mean sizes (2000 and 3610A) shows that although the smaller precipitate size also promotes much higher primary creep rates, the extent of primary creep remains much lower than in the alloy contain 7.5% cobalt (Fig.5).
        4. A further investigation of the role of cobalt shows that the elimination of this element improves the overall creep behaviour in the temperature range 760−1050℃.
        5. The anomalous behaviour of MXON−7.5 Co alloy arises from the extended inhomogeneous {111} <112> slip during the very early stages of primary creep.
        6. mode
        7. primary

Fig 4: 因果関係情報

録に手間がかかり、コンピュータ記憶容量の低廉価と大容量化に伴い、一次情報が増加している。

2.2 学術情報データベース検索に関する問題点

学術情報の検索方法として、最も一般的な方法は、キーワードによる検索であり、自動的あるいは手作業により、得られたキーワードとのマッチング処理を行なう。しかし、近年はコンピュータ能力が向上した結果、フルテキストとのマッチング処理を行なうことが実現可能となった。しかし、いずれも、単純な検索であるため、

1. 検索された情報が多過ぎる（あるいは少な過ぎる）
2. コンピュータパワーが増加したが、フルテキストサーチの場合、検索に時間がかかる。
3. 類似語での検索ができない。
4. キーワードの付け方が問題である。
   著者がつけるキーワードが少ない。
   あるいはキーワードのない論文には、キーワードを付与する必要がある。
5. 検索されるのは論文、文献のような冊子であるが、必要とする情報の基本単位は、必ずしも冊子ではなく、章、節、文単位あるいは個々のデータである。
6. 利用者の検索特性や検索履歴を再利用できない。
7. １つの情報をさまざまな形式に変換して提示することはできない。
8. ある目的のために収集した学術情報をまとめたいが、現在の技術では実現できない。

のような問題がある。

THE DEVELOPMENT AND CHARACTERZATION OF A HIGH PERFORMANCE EXPERIMENTAL SINGLE CRYSTAL SUPERALLOY

	1	2	3	4	5	6	7
creep life	—	—	—	—	—	—	—
creep strain rate	—	—	—	—	—	—	—
creep behaviour	—	—	•	—	—	•	—
creep mode	—	—	—	—	—	—	—
primary creep	—	—	•	—	—	•	—
secondary creep	—	—	•	—	—	•	—
tertiary creep	—	—	—	—	—	—	—
creep strength	—	—	—	—	—	•	—
creep recovery	—	—	—	—	—	—	—
creep rupture	—	—	—	—	—	—	—
creep rupture life	—	—	—	—	—	—	—
creep test	—	—	—	—	—	—	—
creep rate	—	—	•	—	—	•	—

横軸

- 3:heat treatment(熱処理)
- 6: γ'(ガンマ相)
- 15:precipitation(析出)

Table 1: 金属特性・要因間の因果関係の一覧表形式による表現

2.2.1 学術情報データベースの利用実態

これまで、これら学術情報データベースに対し, (1) データベース検索、(2) ユーザーインタフェース、に関する研究が主であった。データベースの利用者が、データベースを検索して得たデータをどのように利用するかに関する研究は少なかった。

一方、企業内部では、検索したデータを有効に利用するために、検索を専門に行なうサーチャと呼ばれる専門職を配置したり、あるいはデータベース代行検索機関に依頼することがある。特に薬品製造、石油・化学、機器製造、食品業界では、(1) 基礎研究着手時の特許調査や重複研究調査、(2) 応用研究段階における競合他社の動向調査、(3) 製品開発段階における市場動向調査などがあり、これらはいずれもサーチャによってなされている。そのような専門職を持たない研究所では、利用者自らこのような作業を行なっている [1]。

いずれにしても、今後、データベースの大量化あるいは分散化が拡大するにつれ、このような知的活動を支援する要求が益々拡大する。

2.3 学術情報データベース利用に関する要求仕様

以上より、学術情報データベース利用者からの要求仕様は、(1) レベル別の要約機能、(2) 柔軟な質問応答機能、(3) 論文内容の分割 及び 斜め読み機能、(4) 複数のアクセスパスによる検索機能、(5) 論文間の散策機能、(6) 複数情報源のダイジェスト版作成機能、(7) 類似研究検索機能、(8) 研究動向提示機能とする、

3　金属材料論文の内部構造分析

金属材料論文を対象として、論文情報ベースを構築するためにその内部構造を分析した。予備調査として、金属材料の専門家が作成していた各論文の要約を参考にし、専門家からの教授を受けた。調査結果は、以下の通りである。

1. 金属材料の論文の多くは、実験を主体にしており、(1)title, (2)authors, (3)affiliation, (5)literature, (6)date, (7)summarizer, (8)objective field, (8)purpose, (9)experimental procedure, (10)experimental result に構造化されていることが分かった。
2. この構造の中では、(9)experimental procedure, (10)experimental result が重要であるため、(9) については、金属材料の名称及び組成、実験前準備、実験名称及び方法に、(10) については、実験結果、考察に細分化した方が望ましい。また、専門用語および英語慣用句とのマッチング処理を使うだけで、ほぼ分類できる見込みが得られた。
3. テキスト情報だけでなく、観察結果や実験結果を示す画像情報として、図あるいは表の情報が重要な要素である。実験結果として定量的に得られたデータを表形式にまとめ、定性的な評価を本文中で行なうことが多い。しかし、表はその形式が不定であるであるため、表中の値を機械的に解析することは困難である。電子顕微鏡などの画像データからその特徴を抽出することも非常に高度な専門的な知識が必要である。
4. 得られた実験結果とそれに対する原因の考察が、その論文の非常に重要情報である。

4　論文内容の効果的提示方法

本研究で、我々が提案する論文情報ベースにより、論文内容を効果的に提示することが可能となる。これにより、2.3 で述べた、学術情報データベース利用に関する要求仕様の中で、(1) レベル別の要約機能、(7) 類似研究検索機能、(8) 研究動向提示機能を実現できる。論文内容の効果的な提示方法として、論文要約機能、複数論文サーベイ機能のそれぞれについて、詳細に検討した。

4.1　論文要約機能

要約方法にはいくつかのレベルが必要であり、最も詳細なレベルすなわちオリジナルな情報である「論文」と最も要約された「論文タイトル」の間でレベル別の要約方法を検討した。また、文章によって表現される「論文」は線形であるが、それを記述する著者の思考は線形ではない。そこで、文章構造にとらわれない要約方法を検討した。

- キーワード
  数語程度のキーワードで論文内容を要約する。自動的にキーワードを抽出する方法としては、(1) 論文内に出現する専門用語の頻度分析による方法、(2) 次に述べる箇条書を要約する方法がある。
- 箇条書 (論文構造に基づく)
  金属材料論文は、３章で述べたように、多くが実験主体の論文であるため、(1) 実験材料、(2) 実験方法、(3) 実験結果、(4) 実験結果に対する考察の構造をしている。従って、この構造に基づいて要約することが必要である。
- 項目間の関連
  金属材料論文は、最終目的は金属の特性を向上することであり、そのために各種の実験を実施している。そこで、(1) 実験処理方法と金属特性の関連、(2) 材料の構造と金属特性の関連、(3) 材料の構造間の関連、などの関連に関する記述が重要であり、これらは一種の因果関係を表現している。これら関連を記述している文章を抽出することが必要である。尚、この項目間の関連は、上記の箇条書 (論文構造に基づく) を詳細に記述したものであり、表現方法としては (1) 一覧表形式、(2) ネットワーク形式が適している。
- 図表情報の要約
  箇条書 (論文構造に基づく) で、各論文内の図表情報の簡略な一覧を提示しており、その一覧を詳細に要約したものである。個々の図あるいは表ごとに、(1) 実験の処理方法、(2) 金属の構造、(3) 実験結果 別に分類して表示する。
- 箇条書 (論文構造に基づかない)
  金属材料に関する研究は、実験が主体であり、試行錯誤の実験を繰り返し行なった上で、論文が作成される。しかし、文章として表現されると、(1) 実験材料、(2) 実験方法、(3) 実験結果、(4) 実験結果に対する考察、の構造で記述されるため、その過程は表現できない。試行錯誤の過程を忠実に再現することは困難であるため、本研究では、論文の構造にとらわれず、金属の特性に着目して、その特性ごとに (1) 実験結果、(2) 実験結果に対する考察の形式で要約することにした。
- マップ形式
  金属材料論文では、実験結果や観察結果を図あるいは表で論文内に添付し、本文中から引用することが多い。１つの図あるいは表が、複数の場所[1]で参照される。「論文」は線形にしかその内容を表現できないため、参照されるつど読者は該当のページに移動する手間が発生する。そこで、マップ形式で表現することで、論文内容を効果的に要約することが可能となる。

4.2　複数論文のサーベイ機能

数１０件あるいは数１００件の論文を、人手でサーベイを行なう時、サーベイすべき論文かどうかの判断したり、自分の研究内容と類似した論文を検索したり、論文の傾向についてその分布状況を調査したりする。従って、複数論文のサーベイ機能として、(1) 論文の傾向の分布、(2) 類似論文の検索、を検討した。これを実現するために、前節で述べた個々の論文の要約である箇条書を使うことにする。個々の論文から得られた因果関係を使った、(3) 複数論文の統合による知識獲得、等についても検討した。

- 構造化された論文タイトルによる分類
  論文タイトルは、論文内容を端的に表現している。金属材料論文の場合、どんな材料を使って、どんな処理方法を行ない、どんな構造の変化があり、どんな金属特性が向上したか、を述べるために、各々の要素の代表的なキーワードで論文タイトルを表現している。従っ

[1] 図表情報は、実験結果の章あるいは考察の章のように複数の場所で参照されることがある。

分類	機能	システム化
要約	論文内容の構造化による要約１	自動化 (Fig1)
	論文内容の構造化による要約２	自動化 (Fig2)
	図表のリンク結合によるマップ表現	半自動 (Fig3)
	因果関係情報	自動化 (Fig4)
	因果関係ネットワーク	手作業 (Fig8)
	特性・要因間の因果関係の一覧表示	自動化 (Table1)
	因果関係の内容一覧	自動化 (Table3)
	図表情報のテーブル表現	自動化
サーベイ	類似度の２次元空間配置	自動化
	類似度の３次元空間配置	自動化 (Fig6)
	類似度の説明機能	自動化 (Table4)
	３次元表示による因果関係の分布	自動化 (Fig7)

Table 2: **METIS** の機能一覧

て、論文タイトルから各々の要素のキーワードを（半）自動的に抽出し、構造化することが可能である。この構造化された論文タイトルで論文を分類できる。

- 論文間類似度による空間配置
  箇条書形式で要約された論文内容を基に、論文間類似度を空間配置表現する。空間上で、近い位置に配置された論文は類似しており、離れた位置に配置された論文は類似していない論文であることを表現できる。従って、この空間配置から類似度についてあいまいな検索が可能となる。
- 因果関係の分布
  金属材料の特性に対する因果関係に着目し、各論文から得られた要約情報を集計することで、因果関係の分布を表示する。これにより、論文の傾向が把握できる。
- 複数論文の統合による知識獲得
  各論文から得られた因果関係ネットワークを統合する、すなわち、個々の因果関係を統合することで、新しい因果関係を導出したり、あるいは因果関係間の矛盾を導出することが可能である。

5　METIS システムについて

本研究では、論文情報ベースを自動的に構築するために、ＫＰという概念を導入した。ＫＰとは、(1) 人工情報系（コンピュータ), (2) 自然情報系（人間）のいずれにも理解可能であり、かつ、弱い構造を持つ知識メディア表現の１つである。

ＫＰを使い、金属材料論文情報ベースを自動的に構築し、その内容を効果的に表示するプロトタイプシステム (METIS システム) を開発した。METIS システムの概要について説明した後、個々の機能について詳細に説明する。

METIS システムは、マークアップ言語で記述した論文を入力とし、(1) 論文情報ベースの構築、(2) 論文情報内容を効果的に表示する。前者は (1) 論文からＫＰ作成部、後者は (2) ＫＰから論文要約作成部, (3) ＫＰから論文サーベイ作成部から構成される (Fig5参照)。機能一覧を Table2に示す。

論文情報の効果的表示機能として、論文要約作成機能と論文サーベイ作成機能を以下に説明する。

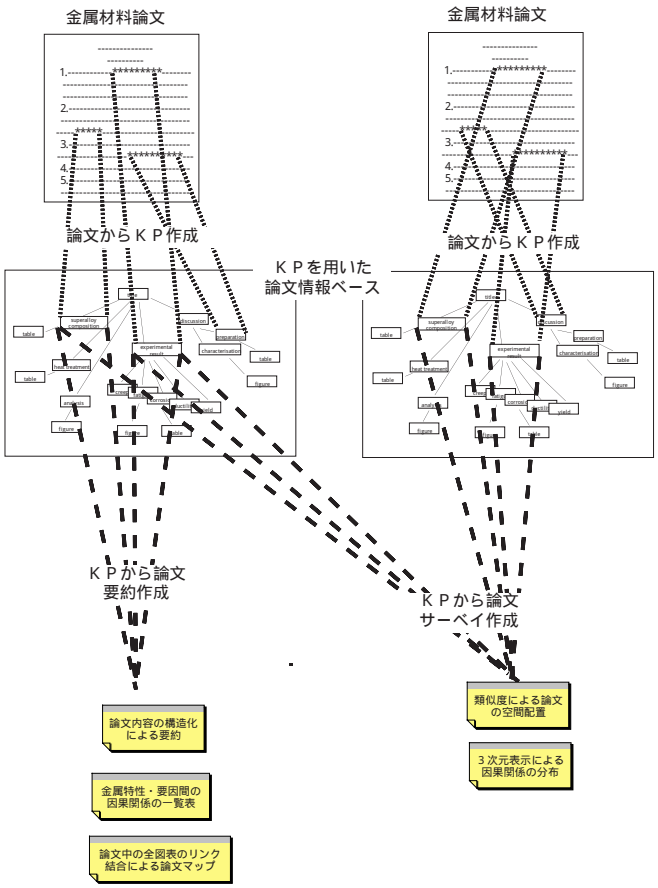


Fig 5: ＭＥＴＩＳシステムの流れ

5.1　ＫＰから論文要約作成

ＫＰを用いた論文情報ベースから、(1) 箇条書、(2) マップ、(3) 一覧表、(4) センテンスの集合、の各形式で論文内容を要約する。この順に、論文内容を詳細に表現する。

1. 箇条書：論文内容の構造化による要約
   要約用のＫＰを使い、２種類の箇条書を作成した。
   - 論文内容の構造化による要約１ (Fig1)
     (1)title, (2)authors, (3)affiliation, (5)literature, (6)date, (7)summarizer, (8)objective field, (8)purpose, (9)experimental procedure, (10)experimental result に構造化されている。この構造の中の特に、(9)experimental procedure, (10)experimental result を中心に、論文内容を要約した。
   - 論文内容の構造化による要約２ (Fig2)
     金属材料論文の研究目的は、金属のある特性を向上することである。特性に着目し、その特性についてどんな実験をおこない、どんな実験結果が得られ、その原因は何であったかという観点から、論文内容を要約した。
2. マップ：論文中の全図表のリンク結合による論文内容のマップ表現
   論文内の図及び表をすべて抽出し、参照関係を用いて、関連あるＫＰから結合する（Fig3参照）。

Effect of processing conditions and heat treatments on mechanical properties of single-crystal superalloy CMSX-2

処理	構造	特性
processed	single crystal	creep behaviour
temperature	porosity	creep strength
process		fatigue behaviour
process		fatigue properties
temperature		low-cycle fatigue
	single-crystal	
	grain boundary	
	single-crystal	fatigue strength
process	microstructure	fatigue strength
processing condition	single-crystal	
heat treatment		
	γ / γ' eutectic	
	dendritic spacing	
heat treatment	μ-phase	
composition	lattice	

Table 3: 因果関係の内容一覧

表示は、ＫＰエディタ[2]で行なうため、ＫＰエディタへの入力ファイルを生成する。
この部分はまだ完全自動化できていない。

3. 一覧表：金属特性・要因間の因果関係の一覧表形式による表現
論文要約 (Fig1参照) では、明示されていない要因と特性の関連を表形式で表現している (Table1参照)。
一覧表の縦軸が金属特性であり、横軸が各々の特性に関して考察している要因である。表形式で表現しているため、どの要因とどの特性について考察しているかを明示できる。現在、 (1) 複雑な因果関係 (例. 中間仮説があるような因果関係) が表現できない, (2) 因果関係の程度や関係の仕方までは抽出されない。
4. 一覧表：因果関係情報の内容一覧
金属材料論文特有の因果関係を示す構造として、ある材料に対して、処理を施した結果、構造が変化し、そのため、特性が変化する。または、処理が直接的な要因となって、特性が変化する。この構造を利用して、因果関係を表現するセンテンスごとに、 (1) 関係情報、 (2) 処理特性、 (3) 構造特性、 (4) 材料特性に関する情報を抽出し、一覧で表示した (Table3参照)。ただし、複数の因果関係を含むセンテンス (複文など) の場合、個々の因果関係がどの項目間について述べているのかは抽出できない。
5. センテンスの集合：因果関係情報
原因と結果の関係を意味する英語の慣用句２５パターンを設定し、それらを含むセンテンスを因果関係情報として抽出し、 (1) 金属材料特性、 (2) 要因別に分類する。 (Fig4参照)

[2] ８章で後述する

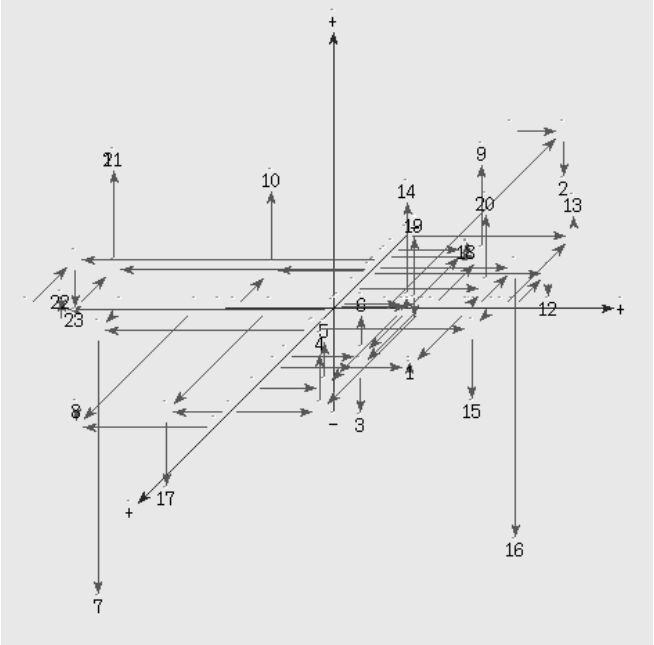


Fig 6: 論文間の類似度の３次元空間配置

5.2 ＫＰから論文サーベイ作成

各論文のＫＰを統合して、 (1) 空間配置, (2) ３次元分布の表現形式で論文サーベイを作成する。

1. 空間配置: 類似度による論文の空間配置
各論文の要約 (Fig1参照) から、各項目 (材料、実験方法、実験結果、考察) 間で類似度を計算した後、それらを総和し各論文間の類似度とする。求めた類似度に多次元尺度構成法を用い、２および３次元上に類似度を空間配置する (Fig6参照)。
空間配置は、明示的に論文の類似度を示すので、類似論文の検索が容易にできる。この空間配置で、利用者が類似していると判断した論文について、その説明機能を提供する (Table4参照) 。利用者が n 個の論文を選択すると、 $\frac{n(n-1)}{2}$ 通りについて、論文間の類似内容を表示する。添付している資料の場合、論文 id が、０３、０４、０５、０６の４論文について類似度を示しており、どの２つの組合せについても類似内容がある。４つの論文すべてについても、 (1) 金属材料が CMSX-2 で、 (2) すべて creep に関する実験を行ない、 (3)creep rate と creep strength に関する実験結果を得て、 (4) 図表情報としてγ', creep, deformation, を参照し、 (5) 実験結果に対する考察としてγ', misfit, single crystal の要因を上げている、という共通点があることが明示している。
2. ３次元分布：因果関係の分布
金属特性・要因間の因果関係の一覧表 (Table1参照) を論文方向に集計する (Fig7参照)。
金属のどんな特性を向上することを目的とした論文が多いかまた、何を要因として考察している論文が多いかを明示的に示す。
軸の説明は、以下の通りである。

similarity of papers among 03 04 05 06

	金属材料	実験結果	考察
03,04	CMSX-2 NASAIR-100	creep rate creep strength	γ' misfit morphology precipitate single crystal
03,05	CMSX-2	creep behaviour primary creep creep rate secondary creep creep strength	γ' heat treatment lattice misfit morphology precipitate single crystal
03,06	CMSX-2 MAR-M200 NASAIR-100	creep life primary creep creep rate creep rupture secondary creep creep strength	γ' lattice misfit porocity precipitate single crystal solidification

Table 4: 類似度の説明

高さ方向は論文数を示し、向かって右向きの軸は金属の特性を、左向きの軸はその特性に対する要因である。金属の特性には、 creep life, creep deformation, creep elongation など４０項目から成り、要因は、処理に関する要因と構造に関する要因に大きく分けられ、前者には、 conventional casting, directional solidification など５項目、後者は、γ phase, γ' phase, single crystal など２３項目からなる。添付している図からは、金属特性の中の creep strength に関する論文で、その要因として、γ'phase, single crystal, heat treatment などをあげている論文が多いことを示している。

6 評価

１０件の論文からヒューリスティックな知識を獲得し、METIS システムを実装した。知識ベースを無修正のまま１５件の論文に適用できた。従って、数１０件あるいは数１００件の論文を入力としても、十分システムの頑強性を示すと思われる。修正が必要な場合でも、ＫＰのマッチング条件部のみの修正で対応できる。

ＫＰからの論文要約の１つに、手作業で行っている因果関係ネットワーク (Fig8参照) がある。因果関係ネットワークとは、一覧表 (Table1参照) では表現できない金属特性・要因間の複雑な因果関係をネットワーク形式で表現したものであり、現在手作業（２ ~ ３時間 / 論文）で作成しているので、省力化する必要がある。

METIS システムについて、金属材料の専門家から、下記の評価を得た。

「実用合金開発の分野では、直接必要な特性のみではなく、極めて多くの因子が関連してくる。これらに関する文献情報を的確に検索するのには長時間の作業が必要で、本研究の因果関係ネットワーク、論文類似度の表示は、文献内容の検索に役立つと思われる。さらに、どのような関係であるかを表示できるようになれば、意義は極めて大きい。論文内容のマップ表現は、理解を助けるのに効果が大きい。さらに工夫してほしい。ただ、不適切な使い方をしている専門用語があるので改善が必要である。金属材料論文は実験主体の論文が多いが、理論やシミュレーションについての論文もあり、これらの論文の要約方法も検討する必要がある。」

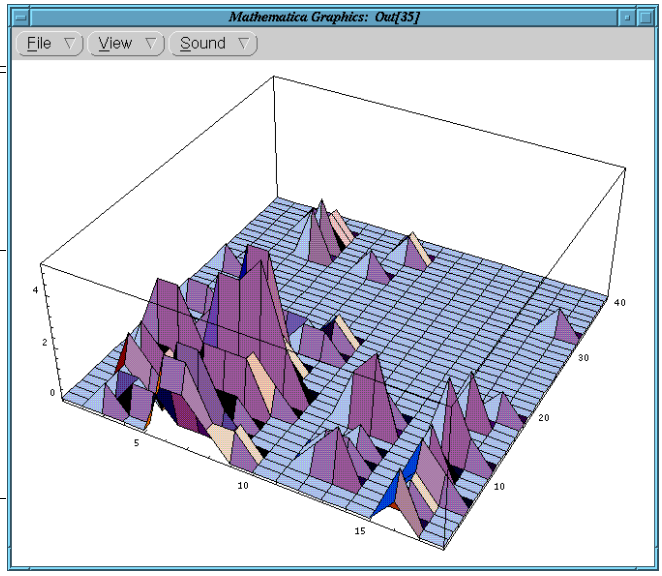


Fig 7: ３次元表示による因果関係の分布

7 まとめ

学術情報データベース利用に関する要求仕様のいくつかを解決するために論文情報ベースを提案した。電子化された論文から論文情報ベースを自動的に構築し、論文内容を効果的に提示するプロトシステムとして、金属材料論文を対象にした、 METIS を開発した。ＫＰという概念を導入することで、論文情報ベースの自動構築が可能になった。
METIS は、論文内容の効果的表示方法として、(1) 論文要約機能、(2) 複数論文サーベイ機能を持つ。１０件の論文で筆者がテストした後、１５件の論文について知識ベースを無修正のままブラインドテストを行ない、金属材料の専門家からの評価を受けた。 METIS の今後の課題について述べる。

8 今後の課題

8.1 ＫＰエディタ

現在の METIS が構築する論文情報ベースは、ＫＰの集合体として表現されているが、利用者とのインタラクションがなく、一方的である。また、ＫＰ作成は浅い自然言語処理を行なっているため、利用者が持っているさまざまな要求や視点をすべて満足できるような論文情報ベースを作成することは困難である。そこで、ＫＰを編集を可能にするＫＰエディタが必要であり、エディタでＫＰを修正することによって、専門家の要求や視点に応じた論文要約やサーベイ作成が可能になる。

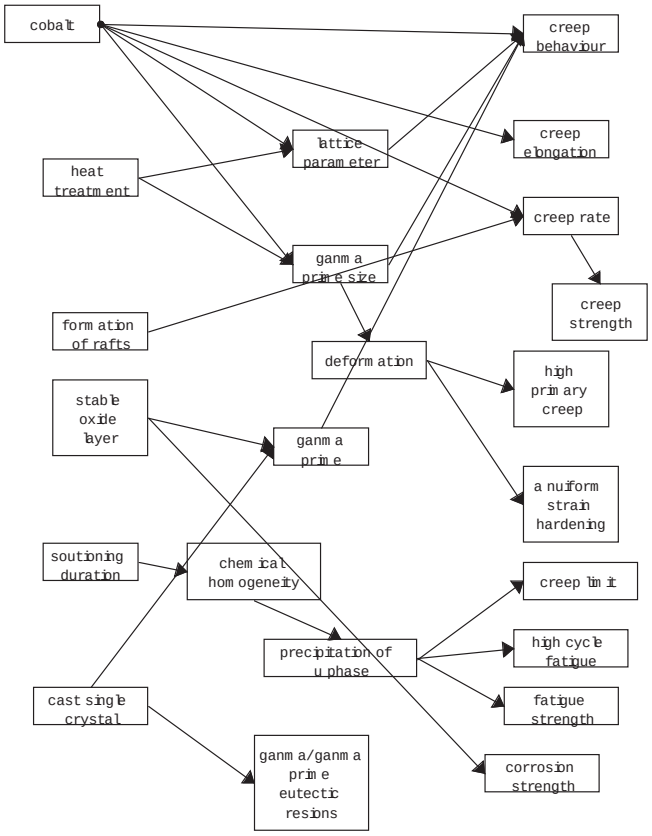


Fig 8: 因果関係ネットワーク

8.2 因果ネットワークの作成支援

因果関係情報内 (Fig4参照) の個々のセンテンスについて、機械的に (1) 関係情報、 (2) 処理特性、 (3) 構造特性、 (4) 材料特性に関する情報を抽出することができた (Table3参照)。

(1) 個々の情報間の関連には、 a. 処理 → 特性、 b. 構造 → 特性、 c. 処理 → 構造、 d. 構造 1 → 構造 2 、の 4 通りしかないことを使い、 (2) センテンス内の詳細な解析をすることで、因果ネットワークが自動作成が可能になる。

8.3 複数論文の知的な統合

現在、 METIS で提供している複数論文のサーベイ機能は、 (1) 論文間の類似度を計算し、その類似度を 2 ないし 3 次元の空間上に配置し、類似度の説明機能を提供したり、 (2) 因果関係の分布を 3 次元で提示する機能を持っているが、不十分であり、知的なサーベイ機能を実現する必要がある。 (1) 論文を自動的にクラスターに分割する機能、 (2) 因果ネットワークを統合し新しい因果関係や論文間の矛盾を導く機能を検討している。

謝辞　金属材料の専門家である金属材料技術研究所の星本健一室長から、専門知識だけでなく、知識メディア化表現に関して有益なコメントをいただきましたこと、感謝致します。

参考文献

[1] 三輪眞木子. データベースサーチャの視点. 情報処理, Vol. 33, No. 10, pp. 1162–1170, 1992.

[2] Mark Stefik. The next knowledge medium. *AI Magagine*, Vol. 7, No. 1, pp. 34–46, 1986.

[3] 寺田浩詔. 英文論文誌の電子化出版について. 電子情報通信学会, Vol. 77, No. 12, pp. 1215–1220, 1994.

著者略歴

松尾 利行
1982 年 京都大学工学部数理工学科卒業、同年住友重機械工業株式会社入社。
1987 年 株式会社日本情報サービス (現 日本総合研究所) 入社。
1994 年 奈良先端科学技術大学院大学情報科学研究科博士前期課程入学、現在に至る。

武田 英明
1986 年 東京大学 工学部卒業。
1988 年 東京大学大学院工学系研究科修士課程修了。
1991 年 東京大学大学院工学系研究科博士課程修了 (吉川・冨山研究室)。
1991 年 財団法人日本システム開発研究所嘱託研究員。
1992 年 ノルウエー工科大学 (Norwegian Institute of Technologoy, Norges Tekniske Hoegskole, NTH) postdoctoral fellow。
1992 年 7 月ノルウエー王立科学技術研究会議 (Royal Norwegian Council for Scientific and Industrial Research, Norges Teknisk-Naturvitenskapelige Forskningsraad, NTNF) research fellow (至 1993 年 9 月)。
1993 年 奈良先端科学技術大学院大学情報科学研究科 助手。
1995 年 同助教授 (知能情報処理学講座)。
東京大学工学博士。

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。