

INFORMATION AND KNOWLEDGE NEWS

情報知識学会ニュースレター

1991 10.1

10

情報知識学会事務局 発行 〒101 東京都千代田区和泉町1番地(凸版印刷㈱内) TEL03(3835)5550 FAX03(3839)6061 ISSN0915 1133

知識の宝庫

コンピュータの世界では、情報の索引や概念別の分類語彙のデータベースを、簡単に”シソーラス”と呼ぶことがあります。この言葉は一般には百科事典や辞書を指します。もとはとえばギリシャ語の $\theta \epsilon \sigma \alpha \nu \rho \omicron \varsigma$ に由来し、”宝の壺や入れ物”を意味します。ギリシャ人が、たった9千人で2万人のベルシャ軍を撃退できたのは、”金”や”権力”によってではなく、優れた知恵(ソフィア)だったことは有名です。彼らにとって知恵こそは愛(フィロ)の対象にふさわしい宝だったのでしょう。

一体誰が蓄積された知識のデータベースをシソーラスと命名したかは知りません。しかしそこには、現代の我々にとっても知識や知恵は宝の山であるという気持ちがこめられていると思われます。今や知識は”情報”としてコンピュータにたくわえられます。さしづめ現代の宝はCD-ROMやフロッピー等に蓄積された電子化辞書や百科事典を始めとする様々なデータベースということになるのでしょうか。

ダイヤやルビーといった普通の宝は希少です。だから人は血眼になって探し求めます。しかし情報という”宝”は、増える一方で減ることはありません。又少数の人に独占されることはなく、皆でシェアすることができます。宝を捜し、掘り起こし、蓄え、加工する作業に加えて、”宝”を”山分け”する技術の需要も高まっています。

本号も多くの方々の御協力で盛り沢山の”宝の山”を満載することが出来ました。今後も”宝の持ち腐れ”にならないよう、どんどん記事を送って下さい。

長瀬

目次

知識の宝庫	1
G・H・ミードのテキスト・データベース	2
ACH・ALLC 国際会議報告	4
データベース構築の現場から	6
TEIの活動と今後の展望	9
紹介「翻訳の為の専門用語マニュアル」	11
電子化辞書ニュース (Ver.2.0,'91)	12
情報知識学会'91 秋季セミナー	13
学会カレンダー (Ver.3.0,'91)	14
情報処理学会「人文科学とコンピュータ」研究会	15
金属材料データベース KIND の構築	16
第3回担当者会議議事録	19
情報知識学会通信	20

G・H・ミードのテキスト・データベース

後藤将之（文部省放送教育開発センター）

筆者は現在、アメリカのプラグマティズムの哲学者・社会心理学者である、ジョージ・ハーバート・ミード(George Herbert Mead, 1863-1931) が残した約70編の学術論文のうち、社会学および社会心理学の領域で、現在でも頻繁に援用される有名論文を中心にした23編の論文を、初出雑誌から採録してテキスト・データベース化する作業を行っている。ここでは、採録にあたっての留意点と、データベース作成の実際に関して説明したい。

(1)採録された論文と採録の理由

人文社会科学系のデータベースでは、原データの選定および採録の方法が、きわめて重大な問題となる。実験などの検証手続きが一定の客観性を保証するように標準化されている自然科学系のデータベースと異なり、人文社会科学系のデータベースでは、どんな資料をどんな形式のデータベースにするかという事が、大きな問題になってこざるをえない。というのも、人文社会科学における主たる方法は、それがどれほど数量化されていても、一定の現象の意味の理解と解釈という作業抜きにはありえない。そして、こうした方法が援用されている限り、何が偏りのないデータであるのかに関して、一意的な定義は、なかなか存在しがたいからである。

こういう理由から、人文社会科学の領域では、資料の選定とそのデータベース化の方法いかんによって、分析結果に大きな相違が生じる。こうした条件下では、資料の選択および採録や分析の方法に関して、明確な表示を行い、どんな目的にとって有効であるかを示しておく必要があるだろう。

本データベースに採録した論文の、選定基準および採録方法は次の通りである。

ミードの著作としてこれまでもっとも有名だったのは、唯一邦訳された書物である『精神・自我・社会』(Mind, Self, and Society, 1932)であった。しかしこの書物は、現実には、生前のミードの自筆になる著書ではなく、シカゴ大学におけるその講義の速記記録をもとにして、弟子のひとりである記号学者チャールズ・モリス(Charles Morris)が編集したものである。長くミードの第1に重要な著書とみなされてきたが、モリスの編集方針などに関して、現在では、いろいろな疑義が提出されている。こうした事情は、ミードの死後、1930年代に相次いで刊行された著書のすべてに妥当する。

このような理由から、ミードが生前に残した雑誌論文を研究する傾向が近年さかんである。本データベースでは、こうした状況に鑑みて、学術雑誌などに掲載された雑誌論文で、比較的言及されることが多いものを、その初出雑誌から採録している。これらの論文群によって、ミードの残した思索のおよその輪郭を知ることができるものと考えられる。

次ページの表1に、こうして採録されたミードの論文のタイトルを一覧表にして示しておこう。

-
- ① The Working Hypothesis in Social Reform, 1899.
 - ② The Definition of the Psychical, 1903.
 - ③ The Relation of Psychology to Philology, 1904.
 - ④ The Basis for a Parents's Association, 1903-4.
 - ⑤ The Imagination in Wundt's Treatment of Myth and Religion, 1906.
 - ⑥ Concerning Animal Perception, 1907.
 - ⑦ The Philosophical Basis of Ethics, 1908.
 - ⑧ Social Psychology as Counterpart to Physiological Psychology, 1909.
 - ⑨ What Social Object Must Psychology Presuppose ?, 1910.
 - ⑩ Social Consciousness and the Consciousness of Meaning, 1910.
 - ⑪ The Mechanism of Social Consciousness, 1912.
 - ⑫ The Social Self, 1913.
 - ⑬ Natural Rights and the Theory of the Political Institution, 1915.
 - ⑭ Josiah Royce : A Personal Impression, 1917.
 - ⑮ A Behavioristic Account of the Significant Symbol, 1922.
 - ⑯ Scientific Method and the Moral Sciences, 1923.
 - ⑰ The Genesis of the Self and Social Control, 1924-25.
 - ⑱ The Objective Reality of Perspectives, 1926.
 - ⑲ The Nature of Aesthetic Experience, 1926.
 - ⑳ Bishop Berkeley and His Message, 1929.
 - ㉑ National-Mindedness and International-Mindedness, 1929.
 - ㉒ Cooley's Contribution to American Social Thought, 1930.
 - ㉓ The Philosophies of Royce, James, and Dewey in Their American Setting, 1930.
-

表1・テキスト・データ化されたミードの論文一覧

(2)データの記録形式

本データベースの記録形式は、固定長80バイトを標準1レコード長として、このうち第1カラムから第16カラムまでを参照部分として、次のような情報が付加してある。

a)当該論文の発表年

b)当該レコードの属する段落番号

c)当該レコードの属する行番号

d)題目、本文、執筆者名・所属、原注、作成者の注、を識別するための記号

これらのデータは、作成者によって新たに付加されている。レコード上の体裁は、できるだけ初出に忠実に行った。ただし、1行の長さは、必ずしも初出雑誌と同一にはなっていない。現在、入力時のミスタイプ訂正やイタリック文字の指定など、細かい部分の整理と統一を行っている途中である。完成のあかつきには、オックスフォードのテキスト・アーカイヴに保存する手続きを取りたいと考えている。

== 国際会議紹介 ==
ACH/ALLC'91

Joint International Conference of the Association for Computers and the Humanities (ACH) and the Association for Literary and Linguistic Computing (ALLC)'91

国文学研究資料館 北村啓子

本年3月17日から21日まで米国アリゾナ州立大学において開催された標記国際会議に参加したので紹介する。

ACH, ALLCとも日本ではあまり馴染みがないが、それぞれ発足して20年近く経つ人文科学系での計算機利用を研究実践している代表的な学会である。1989年からジョイントで国際会議を開催している。今回は約200名の参加者があり、日本からは発表者4名を含む7名が参加した。専門別に見ると、日本からは計算機専門家が4名に対して文学、哲学など人文科学専門家が3名であった。全体では計算機専門家は2~3割程度と少なく、その多くは機械翻訳関係者であった。

今回の重要なトピックスはTEI (Text Encoding Initiative) である。これは、文学、言語学、歴史、哲学などのテキストベースの研究を行う人文科学系において、データのアーカイブ化や流通を目的として、書物を機械可読なテキストとする際の標準を定めようというプロジェクトである。具体的には、どんな時代、ジャンル、言語、計算機、アプリケーションプログラムでも使える一般的なテキストのスキーマを規定しようというのである。最終目標はISO標準にすることである。ACH/ALLC/ACL (the Association for Computational Linguistics) の共同プロジェクトであり、多くのワーキンググループで活動が行われている。すでにガイドラインのドラフトversion1.1を作成しており(1990年秋リリース)、今回の会議ではこのドラフトに対する意見を広く集めてドラフトに盛り込んでいきたいという意図があったようである。

TEIに関しては、ワークショップも開催された。TEIのガイドラインから実際のデータ(参加者が持参)を使った実習まで行われ、盛り沢山の本格的なものとなっていた。また、セッションの中にもTEIの現状報告や将来展望の特別セッションが用意され、多くの聴衆を集めた。

全体のプログラムを紹介する。17日はオープニングセッションとして Xerox PARC 所長 Martin Kay 氏の基調演説と両学会代表者からのあいさつがあった。18日から21日までは、3会場でパラレルに35のセッション、100件近い講演が行われた。特別セッションはTEI以外にも、著作権や辞書、テキスト、ソフトウェアのリソースについて共通の興味ある話題が用意されていた。19日午後にはソフトウェアフェアが開催され、30件近く出展された。

一般講演のセッションを大別すると、

- ・言語学的分析のための理論や技術、意図の分析
- ・内容、筋書き、メタファの理解
- ・コーパスの構築と知識や背景の抽出・語彙、文体、スピーチ、文学等の分析手法やツールや分析結果
- ・辞書や語彙知識ベース構築のための知識、技術、ツール

などがあげられる。この他、言語教育、文体学習、アーカイブと文献検索、マークアップとテキスト編集、複雑なテキストの検索技法、統計的手法、メトリックス、機械翻訳などもあった。参考までにセッションタイトルのリストを最後に付す。

全体に言語学関係が多かったが、言語学や機械翻訳などの対象がテキストやドキュメントのような書かれた文章から談話に移りつつあるという印象を受けた。新しい研究では、内容、粗筋の分析や、メタファの理解、また描かれた社会的背景を読みとる、コーパスから知識を抽出するといった、テキストをアブストラクションする方向に進んでいるようである。ツールとしては、やはりHypertext, Hypermediaのキーワードが至るところで聞かれ、手軽なMacintosh HyperCardを利用している例が目についた。

(この会議はもちろん計算機科学以外の国際会議に参加するのは初めてであり、驚いたことがいくつかあるのでここで紹介する。まず、会議のホスト役からの連絡が全て直接本人から電子メールで送られ、こちらからの連絡にも電子メールが使えて便利であったことである。実は会場に行くまでは、電子メールによる連絡は若手が代行していると信じ込んでいた。次に、会議開催中、計算機ルームの4台のMacintosh IIが参加者に開放され、電子メールの送受信やtalkする光景(中には家族と)が見られ、しかも常に待ち行列ができる程の盛況であったことである。これほど人文科学系研究者の間にも電子メールやtalkが日常的なコミュニケーションの道具として普及しているのだという認識を新たにした。さらに、投稿者は印刷の指示となるタグを書き込んだ原稿を提出し、それを学会で編集印刷して予稿集を作るという体制ができていたことである。このタグは(SGMLだと思われる)、例えば<Title><Author><para><ref><indent>のようなもので、このタグに従って予稿集のレイアウトに合わせて編集される。

最後に特筆しておきたいことは、欧米の大学では既に人文科学系での計算機利用自身が1つの学問として認知されており、Humanities Computingという名前の学科が存在するということである。ここで簡単に紹介したような欧米の現状に比較すると、人文科学系で

の計算機利用(というよりカルチャー)では日本は随分遅れをとっていると実感した。

今回は、1992年4月5日-9日、英国オックスフォード大学で開催の予定である。今後日本からも多くの参加を期待したい。

<参考>セッションタイトルのリスト
Computer Science and Linguistic Analysis
Analysis of Linguistic Intent
The Practicing Humanist and the Computer Expert Systems and the Lexicon
Analysis of Style
Extracting the Social Context of Text
Language Instruction
Understanding Plot
Complex Text Retrieval
Tools for Literary Analysis
Metrics
Variation in Speech
Syntactic Analysis
Psychological Responses to Computer-Mediated Instruction
Machine Translation: East <-> West
Archives and Document Retrieval
Appreciating Monuments of French Literature
Defining a National Literary Corpus
Automating Content Analysis
Managing Lexical Knowledge Bases
Editing and Text Markup
Special Session:
Publishers, Intellectual Property Rights, and the Development of Electronic Text
Status Report on the TEI
Artificial Intelligent Approaches to the Study of Metaphor
Perspectives on the TEI
Statistical Methods in Computational Research
Texts, Concordances and Textual Information
others: Plenary Session, etc.

データベース構築の現場から

(株) 三菱総合研究所 情報サービス部 阿部 明宏

はじめに。(株) 三菱総合研究所では、1984年に国立国会図書館の蔵書のデータベース(JAPAN/MARC)を、実験システムとしてサービスを開始した。サービスシステム名はDIALINEである。

その後、米国商務省のNTIS(National Technical Information Service)データ、日本出版販売(株)の新刊図書データ、日経新聞社の新聞記事情報、第一法規出版(株)の政府政策情報、NHKのNHKテレビニュース原稿情報、東京商工リサーチの企業情報、当社のMRI-REPORTデータなど次々に搭載し、非常に多種多様なデータベースを提供してきた。

当社の調査業務にもよく使われているデータベースである。

DIALINEでサービスしているデータの内容については、東京、大阪で毎年行われているデータベースフェアなどで詳しく紹介しているので、本稿では、データベース構築の立場から、DBMS(Database Management System)を中心に報告する。

DIALINEでは、これらのデータを提供するためのDBMSとして、「BASIS」を使用している。BASISは、米国パテル記念研究所が自社の技術情報の管理、検索を目的として開発された、異色のDBMSである。

どこが異色かと言うと、まず古くから言われている、DBMSのタイプ(階層型、ネットワーク型、リレーショナル型)の、どれにも属さないフラットな転置索引方式のDBM

Sである。また、BASISは、その開発用言語として、科学技術計算用のFORTRAN言語使用され、技術者によって作られたことの一端が窺い知れる。

このようなDBMSは他に例がない。

DBMS「BASIS」のファイル構成は、入力データを保管するQUEUEファイル、データ本体を収容するHEADファイル、転置索引のINDEXファイル、範囲検索を補助するRANGEファイル、シソーラスを収容するTHESAURUSファイル、データベースの構造を定義するテーブルファイル等からなる。

データベースの構築は、このテーブルファイルの作成から始める。テーブルファイルは、DDL(DATA DEFINITION LANGUAGE)言語によって、データベース内のどの項目をキーワードにするかなど、細かく定義する必要がある。

データベース構築のノウハウはここにあり、一度作成したデータベースの構造は、途中で変えることが難しいので、データの特長や、利用者のニーズなど、事前の十分な検討と調査が必要となる。

その点では、最近よく使われているRDBMSの方が簡単で、変更も容易である。RDBMSでは、標準化されたSQLコマンドを使用するが、インデックス以外の項目でも検索ができるので、大型コンピュータから個人用のパソコンまで、機種を問わず普及している。

しかし、このRDBMSにも泣き所がある。それは検索の時間で、検索条件に合致するデ

ータ量に正比例しているため、データ件数が数百件のうちには問題ないが、数千から数万件になると、大型コンピュータでも1-2分かかる事例がある。

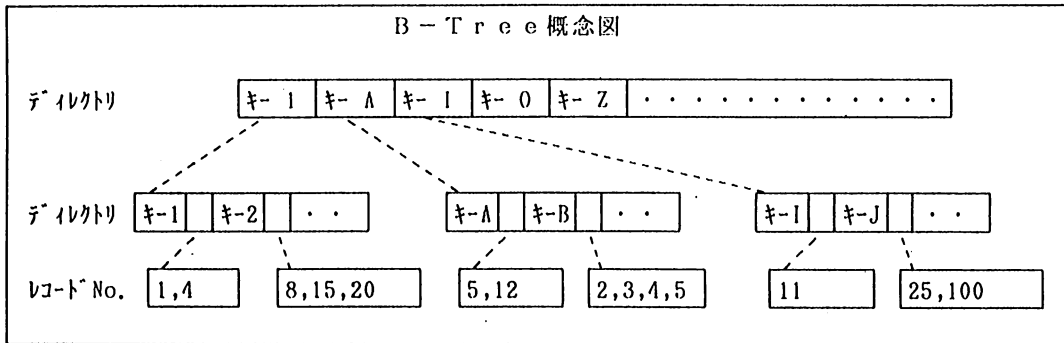
商用のデータベースサービスでは、データベースへ接続している時間で課金するので、検索する件数によって異なるのは、納得出来ないであろう。ちなみに、BASISの検索では、FINDコマンドで検索するが、単一条件の場合は常に3-4秒と一定である。

もちろん、AND検索やAND NOT検索は、検索結果集合の操作なので、集合が大きければそれなりの時間がかかるのはやむを得ないことである。

BASISのINDEXファイルの構造は、B-Treeというファイル形式で、インデックス・ブロックとデータ・ブロックからなっている。

B-Treeは、キーを持つレコードを収めたファイルに対して、特定のキーをもつレコードが入っているかどうかを調べる操作を探索という。

探索を高速に行うには、それぞれのレコードのキーの値とレコードのアドレスとを収めたディレクトリを設け、レコードの件数が多くなれば、ディレクトリを多重にした構成にする。これによって更に高速検索が可能になるのである。(B-Tree概念図参照)



INDEXファイルに密接な関係があるのが、RANGEファイルである。このRANGEファイルは、日付を持ったデータの検索に欠かせない機能で、DATE=910501 TO 910630 (1991年5月1日から1991年6月30日までの新聞記事から探す) というような絞り込みの時に便利である。

もう一つ忘れてならないのが、THESAUURUSファイルである。このファイルは、用語と用語の階層関係(上位および下位)、同義語、関連語を収容し、INDEXファイルと連携して検索を助ける重要なファイルである。

その他の機能に、検索結果集合のSORT

(並べ替え)機能や、検索結果が自由形式で出力できるREPORT機能、一連の検索コマンドを保存しておき、後で何度も実行できるPROFILE機能、検索結果集合内の数値項目の集計ができるCOMPUTATION機能など、非常に豊富な機能を有している。ここまで言っても、まるでBASISの営業マンのようであるが、けっしてBASISに満足しているわけではない。

例えば、DIALOGなどでよく使われている渡り検索(検索の対象を複数のデータベースにする)はできない。また、検索語の統計レポート機能や、何件出力したかなどのログ出力機能もない。

更に欲を言えば、メニュー検索画面のサポートがない。漢字のキーワード検索ができない。しかし、これらは、BASISだけの間

題ではなく、通信プロトコルやパソコン側の検索画面とも深く関わる問題であろう。

データ更新は、追加がほとんどであるが、データベースのサービス時間中にはできないので、専ら夜間に行くことになる。更新の頻度は、日次、週次、月次とさまざまで、その更新時間はデータ量に依存する。毎日データが追加される日経新聞記事データベースの場合、1時間程度である。

ユーザの立場から言えば、365日24時間のサービスが望ましいのだが、構築側から言うと、24時間サービスにするには、同じデータベースをそれぞれ2つ用意し、一方をサービスしている裏でもう一方に追加更新することになる。これは、ディスクスペースが倍必要になり、大きな費用負担になる。

パソコン用のハードディスクは、100メガバイトクラスのものが数10万円で購入できるが、1.2ギガバイトの大型コンピュータ用のディスクは、まだまだ高価である。貴重なディスクの中に、一度も検索されないデータがあるかと思うと、夜も眠れない。

データベース事業は、日夜データを蓄積し続ける装置産業なのである。

システムの維持も重要な仕事である。コンピュータシステムは、実に様々な場面で障害に遭遇する。OS（オペレーティングシステム）のダウン、ディスクの読みとりエラー、通信制御装置の故障、通信回線の故障などきりが無い。

ハードウェアのメンテナンスは、コンピュータセンタの役割であるが、データベースのメンテナンスは、データベース構築者の仕事になる。

DIALINEでは、ディスクの読みとりエラーに備えて、磁気テープへのバックアップ作業を定期的に行っている。ただし、バックアップを取ってあるからといって安心はできない。磁気テープからディスクへ戻す時に、

磁気テープの読みとりエラーが起きるかも知れないからである。

バックアップは、最低でも2重に取っておかねばならない。データベース構築者の心配の種は尽きない。

DIALINEへ接続するためには、最近流行のパソコン通信と同様の準備が必要である。

パソコン、モデム、電話回線、これらをつなぐケーブル、相手と通信するためのプログラム（いわゆる通信ソフトと呼ばれ、市販されている。数千円から10数万円で買える。）が必要である。

揃えたら、通信のために必要な設定を行うが、パソコンから通信ソフトまで正しい設定を行うのは以外に難しい。データベースサービスの窓口では、この対応に追われているのが実状で、DIALINEも例外ではない。

そこで、接続が苦手な人たちのために、DIALINE専用のメニュー方式の通信ソフトの開発を検討している。

パソコン通信サービスが、ビジネス社会にも普及してきた。当社にも、NIFTY-ServeやPC-VANに加入している者が何人かいる。接続料が、10円/分程度なので、個人でも利用可能である。

DIALINEもNIFTY-Serveとのゲートウェイ接続で、更に多くの人々から利用できるようになっている。

今後の課題は、たくさんある。DBMSの項でも述べたが、渡り検索や漢字キーワード検索、検索結果の統計的処理、テキスト情報以外のイメージ情報の取扱い、等々である。

これから越えなければならない山が幾つもあるので、十分な体力（もちろん知力も必要）をつけて、息切れしないように取り組んでいきたいと思っている。

--- T E I の活動と今後の展望 ---

東京女子大学 長瀬真理

現在世界的な規模で文学や哲学等のテキスト・データベースを標準化する動きが活発になっている。その中心になっているのがT E Iである。

日本では文科系の研究へのコンピュータ導入は非常に遅れており、テキスト・データベースの作成もまだまだ欧米に比べると少ない。そのためT E Iの活動はもとより、標準化のガイドラインとして採用されているSGMLについても余り知られていない。

そこでテキスト・データベース研究会(会長:樋口忠治氏、九大)、並びに情報処理学会の部会である「人文科学とコンピュータ」(主査:杉田繁治氏、民博)や電子化辞書研究所(所長:横井俊夫氏)が中心になってT E IのチェアマンであるS. ホッケイ氏を迎え、6月28日と29日にワークショップとシンポジウムを開催した。この機会を利用して簡単にT E Iの活動を紹介したい。

1. 背景

過去30年、西欧では文科系の研究へのコンピュータ利用を促進するために様々なテキスト・データベースが作成されてきた。しかし皆それぞれ独自の方式で機械可読化しており、共通の入力規則はなかった。その為、特定の領域や個人だけに利用されていた。テキスト・データベースのシェア(共有)という観念はなく、仕様等のドキュメントは貧弱であり、たとえ公開されても実際に使用するには多くの困難を伴っていた。一方人文系の研究へのコンピュータの利用は徐々に盛んになり、テキスト・データベースの需要も高まってきた。テキスト・データベースの作成は時間はもとより経費がかかるため、重複作成の

無駄を省くため、公開・共有を目的としたテキスト・データベースの標準的なコーディングの枠組みの作成が急務となってきた。このような状況に鑑みT E Iが発足した。

2. T E Iの概略

T E IはText Encoding Initiativeの略称で、ACH (Association for Computers and the Humanities)、ACL (Association for Computational Linguistics)、ALLC (Association for Literary and Linguistic Computing)の三つの学会のサポートによって運営されている。

基金はUS National Endowment for the HumanitiesおよびEC (European Communities)のDG XIIIおよびAndrew W. Mellon Foundationから出ている。

3. T E Iの活動と目的

T E Iの具体的な目的は:

- 1) 機械可読テキストの交換を可能にする形式を作成すること、
- 2) テキストをエンコードするための要素と要素の表現についてガイドラインを作成すること、
- 3) 主なエンコード法の調査を行いドキュメントを作成すると同時にメタ言語を開発すること、の三つである。

そして具体的には次の3つのコンセンサスのもとに活動が開始した。

- 1) テキストエンコードでの問題点の解明。

- 2) 可能な限り各研究団体・学会の協力を得ること。
- 3) 最近開発されたSGML (ISO 8879規定) をコーディング・スキーマとして採用する。

4. 経過

1988年には、具体的に作業をするワーキンググループをいくつか作り、ガイドラインのドラフトの作成が開始された。

1990年12月初めにOxford大学でミーティングがあり、その時までに各ワーキング・グループはほぼドラフトを完成させた。これが現在入手可能なガイドラインのVer.1である。

1991年11月にVer.2の準備に入る。すでに配られたVer.1からのフィードバックを検討する。

1992年4月のACH&ALLCの合同国際会議時に最終的な討議をして詳細をつめ、1992年夏には最終バージョンを完成させる予定である。

1992年夏に最終ガイドラインの印刷を行い配布を開始する。

現在のところ来年夏以降1年間はECよりファンドがでることになっているが、その後については未定である。

5. TEIと日本

主催者側の一人として筆者は一週間、氏と行動を共にしたが、そのエネルギーな活動には大変驚かされた。上記の2日間のセッションの前に電子化辞書研究所での研究会、東京大学中世英語テキストデータベース・センターでの研究会及び通産省が後援しているSGML懇談会(座長:田中省三氏、富士通)での講演を精力的にこなし、その間”スーパー・アスキー”(記事: '91 Sep.掲載)の取材にも応じるという具合である。

なお日本の文科系研究者側の反応であるが、主催者側の準備不足もあり、特にSGMLについての情報提供が遅れた為、標準化についての議論は余りかみ合わなかった。しかし今

後世界各地で日本の古典のテキスト・データベースの需要が見込まれており、関係する人々にとって国際化という問題は遠からず重要になると予想されることから、熱心な質問が相継いだ。

シンポジウムの参加者は個人的にテキスト・データベースを作成している研究者を始め、テキスト・データベースを作成する主要な機関である、京都大学の人文研、統計数理研究所、東北大学文学部の代表など多彩であった。

一方、既に電子化辞書の作成等で独自の文法に基づくメタ言語を開発している電子化辞書作成所や既にSGML方式でソフト開発等も行っているSGML懇談会での会合では実質的な討議が多く実り多い成果となった。

今後日本側がどのような形でTEIの活動に参加・貢献するかは未定であるが、現在京大の長尾真先生を中心に日本側の受け皿作りが検討されている。日本語の特殊性をうまく生かす形のエンコード法の開発は、ぼちぼち個人的に作成されているテキスト・データベースの共有や作成を促進するためにも必要であろう。あらゆる分野を問わずオープンな形で標準化の問題や国際貢献を協議・検討をする場の設立が急務である。

最後にホッケイ氏の略歴を紹介しておこう。専門はエジプト学で、Atlas研究所を経て、長らくOxford University Computing ServiceのCentre for Teaching InitiativeならびにCentre for Textual Studiesの所長をしてこられた。この秋よりアメリカに設立されるUS National Center for Machine-readable Textsの所長に就任することが決っている。氏は又Computing in the Humanities及びOffice for Humanities Communicationの代表であると同時にALLCのチェアマンでもある。

以上、本学会の学会誌は日本で唯一のSGMLの適用出版物であり、会員の方々にはSGMLに関心の深い方が多い。この機会を利用して、SGMLのアカデミック版ともいえるTEIの活動にも理解を深めて頂ければ幸いである。

紹介：「翻訳の為の専門用語マニュアル」

"Richtlinien für die Konventionelle und Computerstutzte Übersetzungsorientierte Terminologiearbeit"

International Information Center for Terminology, Infoterm

(C) Copyright Infoterm 1991, All Rights Reserved

国際ターミノロジー情報センター (International Information Center for Terminology, Infoterm) では、"Richtlinien für die Konventionelle und Computerstutzte Übersetzungsorientierte Terminologiearbeit"を作成し、このほどその邦訳「翻訳の為の専門用語マニュアル」が完成しました。

このマニュアルは実際の翻訳業務に役立つことを目的にしており、専門用語を体系的にまたは統一的に管理するためのガイドラインです。おもにコンピュータの利用方法について述べていますが、基本的にはカードによる方法にも利用できるようになっており、広く翻訳等の業務に従事する人に参考になると思われます。

各専門分野の研究者は既に専門用語の整理・統一という問題に長年にわたって取り組んでおり、その成果は様々なガイドラインに現れています。しかしここ数年、質の高い翻訳のための専門用語の適切な管理への要望も高まっており、今回はそれに応じるために、翻訳者と専門用語研究者 (terminologist) との協力により作成されました。

目次

- 第1章 はじめに
- 第2章 基本規則
- 第3章 専門用語項目の基本フォーマット規則
- 第4章 データ・カテゴリーとその表記
- 第5章 出典
- 第6章 コンピュータによる専門用語記録
- 第7章 管理システム
- 第8章 専門用語に関する協力

付録1 データ・カテゴリー

付録2 データ記録用紙の例

付録3 参考文献

* マニュアルが必要な方は、MS-DOSでフォーマット済みのフロッピー1枚を下記までお送り下さい。無料でコピーして返送してくれます。なお返信料相当の切手を貼り、返信先を明記した封筒も添えて下さい。

金森 國臣 アイ・エヌ・エス株式会社

〒141 東京都品川区西五反田 7-1-9 H Sビル 3F (TEL: 03(3495)4511)

電子化辞書ニュース (Ver.2.0, '91)

* 9号に掲載された”電子化辞書リスト”のソースをお知らせします。

”E-DICTS”: List of Machine-readable dictionaries compiled from Information Received
from Various sources

COMPILER: Russon Wooldridge <WULFRIC@VM.EPAS.UTORONTO.CA>

VERSION : 5 March, 1991

* 文科系の研究・教育に関連する電子化辞書・ソフト・データベース・研究論文等の網羅的な情報は以下の文献に掲載されています。こちらの方は値段や入手先も詳しく掲載されています。編集委員はText Encoding Initiative (TEI)の主要メンバーで構成されています。現在 Vol.3の発行の準備中です。自薦・推薦を問いません。海外に紹介したい研究論文・ソフト情報などありましたら編集部にご紹介下さい。

The Humanities Computing Yearbook Vol.1 (1988)既刊

The Humanities Computing Yearbook Vol.2 (1990)好評発売中

The Humanities Computing Yearbook Vol.3 (1992)準備中

Ed. by Ian Lancashire and Willard McCarrty

(The Centre for Computing in the Humanities, University of Toronto)

Editorial Board: Susan Hockey

Nancy Ide

Antonio Zampoli

Randall Jones

Donald Walker

Pub. by the Oxford University Press

* なお文科系のコンピュータ利用を促進している国際的な学会の代表的なものにACHとALLCがあります。本号で紹介されたTEIチェアマンのスーザン・ホッケイさんはALLCのチェアマンでもあります。入会ご希望の方は下記にお問い合わせ下さい。又ジャーナルがクォーターリーに発行されています。ALLCは現在研究論文(欧文)を募集しています。こちらも詳細は編集部にお問い合わせ下さい。

Association for Literary and
Linguistic Computing(ALLC)
Dr. Thomas Corns,(ALLC Secretary)
Department of English
University College of North Wales,
Bangor, Gwynedd LL57 2DG, U.K.

Association for computers and
Humanities(ACH)
Joseph Rudman,Treasurer
Department of English
Carnegie Mellon University
Pittsburgh, PA15213, U.S.A.

* 次号では日本で商品化されている電子化辞書(CD-ROM, EB, MT)のリストを紹介する予定です。御期待下さい。

(企画委員：江成保徳、高橋靖明)

〔情報知識学会〕

情報知識学会 '91 秋季セミナー

(主旨)

データ・情報・知識はあらゆる分野の活動を支えるものであります。従来は、自然科学の領域を数多く扱ってきましたが、人文・社会科学の問題を扱うことも極めて重要であり、ますます盛んになると考えられます。このような視点に立ち、最近の新しい試み、あるいは問題点に焦点をあて企画いたしました。情報・知識の諸問題に携わる方々の積極的な参加を願うものであります。

— 記 —

- 日 時： 平成 3 年 11 月 25 日 (月) 13:00~16:30
- 場 所： アルカディア市ヶ谷 (私学会館) 6F 大雪
- 参加費： 会 員
 学生 — 1,000 円、一般 — 2,000 円
 非会員
 学生 — 1,500 円、一般 — 3,000 円
- 定 員： 80 名
- セミナー： 講演 1. 会長挨拶
 13:00~13:20 米田幸夫 (東京大学名誉教授、東海大学教授)
- 2. 文学へのコンピュータ利用
 13:30~14:20 Prof. Armour (慶應義塾大学助教授)
- 3. 考古学とコンピュータ
 14:30~15:20 八重樫純樹 (国立歴史民族博物館
 情報資料研究部助教授)
- 4. 人文科学と情報知識学
 15:30~16:20 藤原 譲 (筑波大学教授)

.....
(申込方法)

- セミナー参加希望者は、下記宛名に郵送あるいは FAX で申込み下さい。

宛名：〒110 東京都千代田区泉町1番地 (凸版印刷 (株))

TEL. 03-3835-5550

FAX. 03-3839-6061

- 定員になり次第締め切らせていただきます。

学会カレンダー (Ver.3.0, '91)

- 情報知識学会後援「歴史研究と電算機利用」会合, 1991, 9,21-22
AIRC'91, Sonoma, Valley, 1991,9,23-25
Quantitative Linguistics Conference(QUALICO), University of Trier, Germany
1991,9,23-27
Computational Intelligence, Milano,1991,9,23-27
国際シンポジウム「コンピュータワールド'91」大阪国際交流センター、1991,9,24-25
The Fourth UNB AI Conference,University of New Brunswick, Canada, 1991, Sep.
Computers and Language 1991: Towards 1992, Sheffield City Polytechnic
1991,9,25-27
7th International Oxford English Dictionary Conference, Oxford
1991, 9.30 - 10.1
IntCAD'91,Columbus, Ohio, 1991, 9.30 -10.3
World Micromouse Finals, Hong Kong,1991, 10. 1-3
光メモリ国際シボジウム、札幌、1991,10,1-4
Applied Linguistics, University of Michigan, 1991, 10,4-6
ACC '91,--Austrarian Computer Conf., Adelaide, Austraria,1991,10,6-10
6th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop,Canada,
1991, 10,6-11
1991 IEEE Workshop on Visual Languages, 神戸、1991, 10, 8-11
情報知識学会「人文科学とコンピュータ」--コンピュータの光と影--、大日本印刷研修会館、
1991,10,14
The Digital Image, Norway, 1991,10,17-19
Internatuna Computer Music Conference, Montreal, 1991, 10, 16-20
人工知能基礎論研究会、C S K 情報教育センター、1991,10,23-25
ヒューマンインターフェースと認知モデル研究会、東京、1991,10.23
インテリジェントシステムシンポジウム、大阪工業大学、1991,10,25-26
17th Annual Minnesota Conference on Language and Linguistics, 1991,25-26
ASIS '91,Washington D.C., 1991,10,27
The 3rd CIMI, Santa Monica, 1991,11, 3-4
ISS '91--Int'l.Symposium on Supercomputing, 福岡、1991,11,6-8
Tools for Artificial Intelligence, 1991,11, 5-8
TEI Joint Meeting, Norway, 1991, 11, 15-18
第8回色彩工学コンファレンス、科学技術館・サイエンスホール、1991,11,18-19
3rd Conference on Situation and its Application, Oiso Prince, Kanagawa
1991,11,18-21
グラフィックスとCAD、機械振興会館、1991,11,20-21
情報知識学会'91 秋季セミナー、アウカディア市ヶ谷(私学会館)、1991,11,25
第8回 トロン・プロジェクト国際シンポジウム、TEPIAホール、1991、11,26-27
知識のリフォーマション・シンポジウム、東京大学山上会館、1991,11,27-28
情報知識学会「人文科学とコンピュータ」-- ミュージウム特集--、川崎市市民ミュージウム、
1991,11,29

情報処理学会「人文科学とコンピュータ」研究会開催

第11回研究会 【特集：コンピュータの光と影】

10月14日(月) 10:00-15:30、大日本印刷(株)研修会館(東京市ヶ谷)

10:00-12:00

読書とコンピュータ(当山:円満寺)

聖書とコンピュータ(伊東:岡崎学園短大)

コンピュータは教育に何をもたらしたか(上窪:NEC)

12:00-13:30

昼食

13:30-15:30

クリエイティブな作業におけるコンピュータ(佐藤:フジテレビ)

芸術とテクノロジー(土佐:武蔵野美大、総合電子専門学校)

情報の生態学(長尾:京大)

第12回研究会 【ミュージアム特集】

11月29日(金) 10:00-17:00、川崎市市民ミュージアム(神奈川県川崎市)

市民ミュージアム館長挨拶

10:00-11:00「ミュージアムにおけるコンピュータ利用の動向」

AVCC(榎井:AVCC)、他

11:00-12:30「研究発表:技術紹介」

博物館におけるマルチメディアDBネットワークの構成法(有泉:九州NES、打浪:九工大)

INSネットを用いた高精細画像ネットワークシステムの構成法(堀口:NTTデータ)

CG画像のホログラフ化と博物館への応用(大越:大日本印刷)

12:30-14:30

昼食

14:30-17:00「事例報告と討論」

徳島県立近代美術館(吉原:県立西洋美術館)

東京国際美術館(遠藤:東芝)

根津美術館(西田:根津美術館)

諏訪市立博物館(高見:諏訪教育委員会)

神奈川県立-千葉県立中央博物館(大場:千葉県立中央博物館)

川崎市市民ミュージアム(濱崎:川崎市市民ミュージアム)

博物館と情報(中山:筑波大)

17:00-17:30「市民ミュージアム施設見学」

*いずれも入場無料。プログラムは多少変更される予定。詳細は下記へお問い合わせ下さい。

洪 政国(日本IBM東京基礎研究所、Tel:03(3288)8280、Fax:03(3265)4251)

『金属材料データベース KIND の構築』

東北大学金属材料研究所 川添 良幸

1. 序

材料関係の研究は、従来実験を主体として行われてきた。理論的な解析ももちろん精力的に行われては来たが、本来複雑な対象を扱う学問であり、新たな機能を具備した材料の創成まで可能な理論的枠組みの構築は、ほぼ不可能であると言わざるをえなかった。

一方、材料に対する機能性高度化の要求が高まり、それに対処するため材料の組成はさらに複雑化しつつあり、限られた研究者・設備で迅速に有効な結果を得ることが困難になった。また、日々蓄積されつつある膨大な研究成果を限なく検索することも旧来の方法では極めて困難になっている。

こうして、材料に関するデータベース構築、およびそれを活用したコンピューターによる材料設計システムの実現は世界的な急務となってきた。

2. 材料データベース

INSPEC等の商用データベースは、数ヵ月遅れで科学技術論文のほとんどの分野を網羅して、題名、著者名、雑誌名、アブストラクト等を提供する。しかし、これらのデータベースには、我が国で作成されているものは少なく、圧倒的に輸入超過であり、新たな貿易摩擦の原因になりかねない。今後、我が国が先進国と呼ばれ続けるためには、独自のデータベースを構築し、諸外国にもそれを提供して行かねばならない。

現在、材料全般を網羅したファクト・データベースはなく、各対象ごとに作成されたものが統一されずに存在している。それらは、熱力学データや相平衡データ等の物性別、耐熱材料や耐食材料等の機能別、と分類できる。コンピュ

ーター・データとして入手可能なもの、および本の形で出版されているものがある。

一方、ケンブリッジ・データベースに代表されるように、化学の世界はデータベースなしに語れないという状態になっている。これは、特に薬学でドラッグ・デザインと呼ばれるコンピューターによる薬物設計が盛んになったことを反映している。スーパーコンピューターの重要な納入先が薬品会社であることも、よく知られている。

文科系のデータベースは、文章そのものを入れたテキスト・データベースが主流である。フランスでは出版された全ての本の文章をデータベース化しようとしているほどである。最近、アメリカおよびイギリスを中心としてSGMLによるテキスト・データベースの標準化運動(Text Encoding Initiative)が進められており、我が国でもテキスト・データベース研究会を通して日本語処理部分等に対応している。こうして蓄積されたデータベースは、通常文献の比較・対照、インデックス作成等に活用される。また、諸外国では盛んに研究されているにもかかわらず、我が国にはほとんど研究者のいない計量文献学では、文章そのものを統計処理する。その結果から、例えば客観的な著者推定を行う(宇治十帖は紫式部作か?)。

データを集積した大きなコンピューター・ファイルは、フラット・ファイルと呼ばれ、データベースではない。きちんとした管理システム下に構築され、運用されている状態になって始めてデータベースと認定される。データベース管理システムには、階層型とリレーショナル型がある。階層型は、検索速度が早く大規模なデータベース向きであるが、構造の変更が大変である。それに対してリレーショナル型は、コンピューター・パワーを必要とするが、種々の変更要求に迅速に対応でき、さらに表形式なので

理解が容易である。また、SQLが国際規格として承認され、互換性も高くなった。材料研究では、実験手法の改善等に伴い初期に予期しにくい項目の追加が必要となることが多く、それに対処するため今後リレーショナル型のデータベースが活用される機会が増えるものと考えられる。

内容のあるデータベースは、良い管理システムのみから生まれる訳ではない。意味のないデータや間違った値を蓄積することは、時として何もないより悪い状態を招来しかねない。データベース構築は、その目的・内容を十分に理解した熟練者によってなされなければならない。単なる人手ではなく、真に頭脳・経費の必要な仕事である。利用したい人は多く、構築にまわる人が少ないのは、基本的に縁の下の力持ち的な業務だからである。

コンピューターのネットワーク化が進み、データベースも世界的な広がりやオンライン検索できるようになってきた。その理由の一つとして、通信のコストが下がったことが挙げられる。各種のデータベースを自在に検索できれば、いながらにして世界の最新情勢を入手できる。しかし、多岐に渡るデータベース中から必要とする適切なデータを抽出することは極めて困難である。そのため、専門職としてのデータベース・サーチャーが重宝がられている。旧来の図書館司書の資格ではなく、より現代的な要望に適合した新種の職業の誕生である。

こうして構築した材料データベースを人間が検索し、従来の方法で解析に利用している範囲では、宝の持ち腐れである。問題に対応した優れたエキスパート・システムによって効率的な検索を行い、必要な情報を迅速に抽出し、材料設計を行うことが望まれる。対象を限れば、実用段階に達しているシステムも存在しているが、一般的には今後の重要な研究課題である。

ドラッグ・デザインのソフトウェアは、全て高度なグラフィックス・インターフェースを伴っている。研究者がイメージを膨らませるためには、材料設計システムにおいても画像ではなくてはならない存在である。もちろん、初期形状

の入力も、数値としてではなく図形インターフェースを介することが必須である。さらに、動的なシミュレーションには、アニメーションも欠かせないテクニックである。

3. 実例

以上、コンピューターによる材料設計およびそのための基礎資料となる材料データベース構築についての概要を述べたが、ここで事例を紹介してより理解を深めてもらおう。具体的には、我々の合金設計制御工学研究部門の研究テーマを挙げ、その概要を示す。

東北大学金属材料研究所は今年で創立75周年を迎える、26部門からなる国立大学随一の規模を誇る全国共同利用材料科学研究所である。創立者本多光太郎先生以来、実験を主とした伝統を堅持し、数々の目覚ましい成果を上げてきている。最近、英語名称を Institute for Materials Research と変更し、金属に止まらず各種構造および機能材料までその研究対象を広げ、さらに精力的に研究を行っている。

一方、この十年間、本研究所は複雑化する一方の材料研究に必須となりつつある材料データベースの構築にも積極的に取り組んできた。一例として、東北大学大型計算機センターで国内の大学の研究者にサービス（利用頻度一位）している金属材料データベース METADEx は本所が購入、提供しているものである。またこの3年間は文献一次情報の収集に当たり、現在約2万件の資料を光磁気ディスク HITFILE 上に蓄積し、KINDと名付けている。同時にこれらの文献からキーワードを抽出して二次情報を作成しており、その部分はやはり大型計算機センター経由で全国共同利用に供している。

また、昨年度は研究所内の情報関係担当部門として材料科学情報室を設置し、所内に新電話・ファックス系を導入し、学内コンピューター・ネットワーク TAINS および材料科学情報室の IBM9370 上の SQL データベース・システムとの連携により各研究室から一次情報を検索・出力可能とした。このような大規模な

システムは大学関係では諸外国も含め未だ例を見ないものである。さらに、現在より大規模なコンピューター・システムの導入を計画している。

4. まとめと展望

複雑化する一方の材料設計において、蓄積された資料の網羅的で迅速な検索は日増しに重要となってきた。本研究所では、光磁気ディスク、データベース管理システム、電話系、およびコンピューター・ネットワークを統合化し、

蓄積中の一次文献データの効率的な検索システムを実現した。今後、こうして集積した文献から有用な数値情報を抽出してファクト・データベースを構築しなければならない。そのためには大量の文献中からの数値データ抽出を自動化する必要がある。我々はそのための基礎的研究を画像認識技術を駆使して行おうとしている。最初のテーマは磁性体を対象とし、磁化曲線、相図等からのデータ自動抽出を研究している。さらに、構築したデータベースを活用して材料設計エキスパート・システムの開発を行う予定である。

編集部へ：Letters to Editors

毎号楽しみに読んでいます。7号より装丁が替わり執筆者の写真等がなくなり残念です。出来れば継続的な連載記事など充実した内容のものも欲しい。(横浜：米原)

編集部から：

本号は編集長が出張のため原稿の締め切りが早まり皆様には大変ご迷惑をおかけいたしました。深くお詫び申し上げます。次号はこれまで通り発行月の前月の20日(11月20日)となります。

なお編集部では様々な情報知識関係のプロジェクトの活動を始め、各種研究会、セミナーの案内・報告を掲載しております。ご協力をお願い致します。

又現在環境問題、あるいはバイオテクノロジー、生命科学等のプロジェクトや研究会の案内をして欲しいとの要望が編集部に届いております。関係者の方の情報提供をお待ちしています。

それから御要望に答え読者の声の欄を設けました。気楽に御投稿下さい。今回は米原様より貴重な御意見を頂戴いたしました。

第3回担当者会議（仮称）議事録

日時 : 平成3年7月15日(月) 18:00~21:00
場所 : 東京大学山上会館 203号会議室
出席者 : 米田、月見里、藤原、安澤、江成、高橋、芦崎、石塚、田隅、長瀬

配布資料: 1 第3回担当者会議(仮称)議事録(案)
2 企画委員会 検討報告
3 情報知識学会誌受け払い記録

報告及び討議:

- 1) CODATA部会の活動について田隅会長より報告があった。
 - ・現在タスク・グループがアジア・オセアニア10カ国のデータを調査中(所在と内容紹介: CODATAの特別報告に掲載予定)
 - ・既に小会議を5回、東京、北京、京都、ソウルで開催。Proc.を東京と北京大会では刊行。改訂版のDB化と出版を計画中。(岩田担当。約100万円の予算)
 - ・部会員の名簿を次田中心に作成。会費納入は43名。
 - ・CODATA本部との連絡は良好。上納金も支払い済み。
 - ・第5回CODATA国際会議の準備中。
(日中セミナー、日本の情報関係の展示を行う予定)
- 2) 長瀬委員よりニュースレター配布の報告。8月発行の9号の準備。
- 3) 江成委員より企画担当の報告。
 - ・人文系をターゲットに11月セミナー、またはワークショップを計画。
(日本史ハイパー辞典の紹介)
(学術会議の依頼により歴史情報資料センター設立の動き)
 - ・詳細は7月25日私学会館で安澤、長瀬と打ち合せ。
- 4) 石塚委員より情報知識学会誌受け払い記録についての報告。
 - ・SGML懇談会で発表した。SGML適用出版物は珍しく、反響大であった。
- 5) 芦崎委員より岩田より英語の論文受領の報告。締め切りを2カ月延期しニュースに掲載。
- 6) 藤原副会長より個人会員&法人会員のupdateについて報告。
- 7) 月見里副委員長より事務所移転について報告。
 - ・正式に事務所移転が決定。ニュースレター9号で案内。庶務・会計の引き継ぎを開始。
次回は9月3日(火)14:00~18:00より凸版で開催。

情報知識学会通信

情報知識学会に入会をご希望の方は、以下のフォームをコピーして必要事項を御記入の上、事務局に郵送、又はFaxで御送り下さい。折返し入会案内、入会申し込み書等の書類をお送り致します。（現在入会金は1,000円、年会費は5,000円です。）なお現在ニューズレターがあります。ご希望の方はお知らせ下さい。

宛先

Fax：03（3839）6061

〒101 東京都千代田区和泉1番地（凸版印刷内）

情報知識学会事務局

担当 齊藤 行

情報知識学会に入会したいので必要な書類を御送り下さい。

個人用

法人用

（どちらかを丸で囲んで下さい）

住所：〒

（フリガナ）

氏名：

電話：

Fax：

学会への御質問あるいは連絡事項：