JSTの英文データベースとインターネットによる提供

高木 津子¹, 加藤秀樹², 菊地博道³

著者抄録 : 科学技術振興事業団 (JST) では, 我が国で発生する科学技術情報を海外に提供するために, 国内で発行された科学技術・医学文献を対象として, 英文のデータベース (JICST-E) の作成し, JOIS及びSTN-Internationalで提供してきた。2000年12月から, 索引付けをする前のJICST-Eのデータをデータベース化し, 新たにWebにより無料で提供を開始した。本文では, 英文データベースの概要, 新サービス概要及び従来のサービスとの相違点を紹介する。

キーワード : 英文データベース, 研究情報, インターネット, Web, 日英機械翻訳システム, J-EAST, 科学技術会議, 科学技術基本計画

JST English text databases and the provision on the Internet

TAKAGI Natsuko¹, KATO Hideki², KIKUCHI Hiromichi³

Author Abstract: JST has long produced a comprehensive bibliographic database with English citation, abstract and index (called JICST-E) covering the literature published in Japan on all fields of science, technology and medicine, and provided it for the overseas users through JOIS as well as STN-International. Since December, 2000 JST has used the JICST-E data remained not indexed yet to make the database and got it available on the Web for the public free of charge. This paper describes outline of the English databases and differences between new service and service we have provided so far.

Key words: English database, research information, Internet, Web, Japanese to English machine translation system, J-EAST, Council for Science and Technology, Science and Technology Basic Plan

(情報管理 44 (3): 190-198)

1. はじめに

科学技術振興事業団では, 我が国で発生する科学技術情報を海外に提供するために, 1985年1月 (当時は日本科学技術情報センター: JICST) より, 国

内の科学技術・医学文献を対象として, 英文データベース (JICST-E) ファイルの作成に着手し, 1986年10月からJOIS, さらに, 1987年10月からSTN-Internationalで提供を開始した。

1994年には, これまでJICST-Eファイルの収録対象としていなかった予稿集などの書誌情報を含むJICST-E Plusファイルを作成し, 1995年5月からSTN-Internationalで提供を開始した。

その後, インターネットの普及など情報提供手段の多様化に対応するため, 2000年12月から, 新たにWebによる提供を開始した。

2. 英文データベースの作成

日本の科学技術情報の入超に対応するため, 作成を始めた英文データベースは, 当初からいくつかの問題が内在しており, これらの問題を解決しつつ現

¹科学技術振興事業団 国際室 (〒332-0012 埼玉県川口市本町4-1-8)

¹Japan Science and Technology Corporation (JST), (1-8, Honmachi 4-chome, Kawaguchi-shi, Saitama, 332-0012)

²科学技術振興事業団 製品管理部 (〒102-0081 東京都千代田区四番町5-3)

²Japan Science and Technology Corporation (JST), (5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081)

³科学技術振興事業団 総務部 (〒332-0012 埼玉県川口市本町4-1-8)

³Japan Science and Technology Corporation (JST), (1-8, Honmachi 4-chome, Kawaguchi-shi, Saitama, 332-0012)

在に至っている。

データベース作成件数, 直接入力(原文献に英文標題が付与されているもの, 原文献に英文抄録が付与されていて利用許諾を受けているもの) による処理件数, 人手翻訳・機械翻訳による処理件数, 抄録付与件数の推移を表1に示す。

2.1 作成経費・処理期間

日本で発生する科学技術情報の多く(JICSTファイルに含まれている国内文献のうち1999年発行の資料によると約8割) は日本語で書かれているため, JICST-Eファイルを作成するためには, 翻訳する必要がある。特に, 検索した際に原文献の内容を把握するために重要となる抄録については, 通常のJICSTファイルで英語文献を日本語抄録として作成する委託費に比べて2倍程度の経費が必要となる。

また, JICST-Eファイルは, 直接入力を除いて, JICSTファイルの作成後に日本語の標題と抄録を翻訳し, キーワードを英語に変換してファイルを作成するため, JICSTファイルの提供より1~2か月遅くなり, 雑誌の発行から見ると半年ほどのタイムラグが生じている。

現在では, JICSTが開発した日英機械翻訳システムを利用して翻訳を行っており, 経費の低減, 期間の短縮を図っている。特に, 導入当初は日本語の文書を機械翻訳システムが判断できるよう前処理したあと, 翻訳処理し, さらに翻訳不能箇所や誤訳などを修正する後処理という行程であったが, システムの改善などにより前処理の行程を省略することが可能となり, いっそうの低コスト化が図られている。

さらに, 作成当初から原文献に付与されている英文の標題や抄録を利用の許諾を得て入力すること

により, 経費の削減および処理期間の短縮を図っている。

2.2 抄録付与率

前述のとおり, 作成経費や処理期間の問題から, すべてのデータに抄録を付けることができない。そのため, 検索の際の網羅性が低いこと(抄録文中の単語も検索語となるため), 検索した記事の内容を判断するための情報が少ないことがある。年々抄録付与率は高めており, 1995年には約50%となったが, 1996年から書誌情報のみのデータを加えたため, 件数に大きな変化はないものの抄録付与率は30~40%と低くなっている。なお, 2000年からは, 抄録のないデータは作成しないこととしたため, データの件数は大幅に減少している。

2.3 日本語の特殊性

原文が日本語の文献が多いため, 検索した結果, 必要な文献を入手しても読むことができる人が少ないなどの理由から, 二次情報として十分に活用できないことである。

3. インターネットによる提供の背景

情報処理技術の発展を踏まえた国の科学技術情報流通にかかわる施策として, ネットワーク時代を考慮した現状認識, 施策推進の考え方, 情報科学技術の戦略的な推進のための共通的な方策に関して, 科学技術会議諮問第25号「未来を拓く情報科学技術の戦略的な推進方策の在り方について」に対する答申(平成11年6月2日) および同第26号「科学技術基本計画について」(平成12年12月26日, 平成13年3月30日閣議決定) には次のように記載されている。

表1 英文データベースの作成件数の推移

	1987	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
直接入力	117,233	160,256	130,524	155,907	144,870	165,017	182,647	248,520	248,656	250,888	212,533	124,707
人手翻訳	68,568	49,488										
機械翻訳		24,414	102,091	86,948	83,621	80,672	91,712	166,352	156,969	141,685	133,632	20,466
抄録付与	50,076	87,035	76,104	95,790	95,787	135,579	137,580	135,140	132,699	128,920	131,159	107,133
合 計	185,791	234,158	232,615	242,855	228,491	245,689	274,359	414,872	405,625	392,573	346,165	145,153

科学技術会議諮問第25号より抜粋

科学技術情報の円滑な流通促進にかかわる課題

我が国の研究開発活動は、その研究成果を世界に向けて広く発信することにより、初めて、世界に認められ、また、アジア地域の中でも我が国が中心的な役割を果たすことができると考えられる。しかし、我が国では、国際的に通用する学協会誌の発行や研究者自身の発信能力に十分ではない面がある。また、所在情報の不足、言語や制度の壁等により、研究コミュニティが発信した情報が海外から十分に利用される状況になっていない。

情報科学技術の戦略的な推進の考え方

「世界に向けた科学技術情報の発信」の中に、「国内で発生する論文、研究報告等について、国際的な流通が円滑に行われるよう、総合的なデータベースとして整備し、英文化した上で海外に発信する必要がある。」

情報科学技術の戦略的な推進のための共通的な方策

「国際的展開」の中で、「科学技術情報の流通に係わる国際協力については、従来、我が国は先進国から活発に情報を入手することが中心であったが、今後は、先進国との間では受信と発信がとりあうよう相互交流を行い発展途上国に対してはこちらから積極的に提供していくという考え方に立って、前述したように、世界に向けた科学技術情報の発信を強化していく必要がある。」

科学技術基本計画より抜粋

国際的な情報発信力の強化

我が国の科学技術活動が国際的に認知され、評価され、その結果、世界一流の人材や最新の情報が我が国に集結するようになるためには、研究成果、研究者、研究機関に関する情報の積極的な海外への発信が重要であり、研究成果の英語での発表を強化するための支援を行うとともに、学協会とも連携しつつ、国際的な水準の論文誌の刊行等、情報の組織的な発信を行うための環境を整備する。

これら政府の方針を受ける形で、JSTは科学技術研究の成果や科学技術政策に関する情報などの所在、概要の案内、研究者の紹介、研究機関の概要などを英語で積極的に発信していくことにより、世界各国と研究情報流通において協力を推進するための1つとしてこれまで作成してきたJICST-Eをベースに、わが国が強い技術分野を中心に研究情報英文化事業を進めることとなった。

4. J-EAST (Jst English Articles of Science and Technology) の提供システムについて

4.1 システム構成

J-EASTシステムの概要を図1に示す。
ソフト構成は次のとおりである。

図1 J-EASTシステムの概念

- ・ OS Solaris7
- ・ WEBサーバ InfoProviderPro
- ・ 検索ソフト Intelligent Search

J-EAST提供システム開発にあたっては、JOISやSTN InternationalなどJSTのWeb上でサービスを行っている既存のシステムを参考にした。また、システムの管理にはできるかぎり省力化・自動化に配慮した。

Web用ソフトにはInfoProviderProを、検索用ソフトには全文検索できるIntelligent Searchを使用している。J-EASTシステムは、検索語に対する絞り込みを行う検索機能や文献標題の結果一覧や文献の詳細表示する機能など、サービス系の機能と、検索を行ったユーザのアクセス管理やデータの更新など、管理系の機能を有している。

4.2 データ項目

記事の収録対象期間は直近の2年としている。データ更新は月1回、平成13年5月の時点で約52万件のデータを収録している。データの検索事項、収録データ項目を表2に示す。

4.3 検索機能

検索においては以下の機能を備えている。

- 複数検索語によるand/or検索
- 発行年の範囲検索

表2 J-EASTの検索項目・収録データ項目

検索条件入力画面	収録データ項目	検索結果詳細画面
検索項目		表示項目
—	整理番号	Accession Number
keyword(KW)	英文抄録	Abstract
	Title(TI)	Title
	Journal Title(JT)	Source
ISSN(ISSN)	ISSN	
CODEN(CODEN)	CODEN	
ISBN(ISBN)	ISBN	
—	JST資料番号	
	レポート番号	
	巻号	
	特殊号	
	ページ	
Publication Year	発行年	
Author(AU)	著者名	Author
	団体著者名	
Corporate Source(CS)	機関名	
Language	原文言語	Language
—	発行国名	PUB.Country
	写図表参	Figure&Table&Reference

原文使用言語による絞り込み検索

検索結果の一覧表示は検索語の出現率の計算を全文検索ソフトが動的に行い、順序を決定する機能

1,000件までの検索結果の一覧表示

4万件以上の検索結果に対する検索絞り込みメッセージ表示

5. 検索画面の説明について

画面の遷移を図2に示す。URLは、http://j-east.tokyo.jst.go.jp/である。またJSTホームページ英語版からもリンクが貼られている。

トップページ(図3)では、本システムの説明文、検索画面へ遷移する「START」ボタン、ヘルプ画面へ遷移する「USER GUIDE」およびJSTホームページへのリンクがある。ヘルプ画面ではJ-EASTの画面の遷移、各画面の説明と検索方法を英語で記載した。システムに搭載されている記事の件数と更新日がここに表示される。

トップページの「START」ボタンをクリックすると検索条件入力画面(図4)に遷移する。検索項目をプルダウンで選択し、対応する検索語を入力する。条件を入力した後に「SEARCH」ボタンをクリックすると、検索結果の一覧が画面下部に表示される(図5)。「CLEAR」ボタンをクリックすると、入力した条件がすべて消えて初期状態に戻る。

検索結果一覧表示画面では、入力した条件とその検索結果のうち、検索語として入力した用語が多く含まれているデータを多いデータから順番に10件の文献標題の一覧、「DISPLAY」ボタンが表示される。一覧の上部に検索された記事件数と一覧のページ、「ALL」ボタン、「CLEAR」ボタン、各標題の左側に連番とチェックボックスがある。ページは当該ページと2ページ先までを指定することが可能で、3ページ前は「Prev.」、3ページ後は「Next」を指定することで、それぞれにページ送りできる機能である。「ALL」ボタンは表示されている文献標題のすべてのチェックボックスを選択し、「CLEAR」ボタンはチェックボックスの選択を外す。検索結果の詳細画面へはチェックボックスを選択し「DISPLAY」ボタンを押すか、直接その文献標題をクリックすることにより遷移する(図6)。「BACK」ボタンには一覧表を取り消す機能がある。

検索結果詳細表示画面の「CLOSE」ボタンを押すと、検索結果一覧表示画面へ戻る。

図2 J-EASTの画面の遷移

図3 トップページ

図4 検索条件入力画面

図5 検索結果一覧表示画面

J-EAST Database	
CLOSE	
COPYRIGHT 2001 JST	
Accession number	00A0589557
Title	Structure and Function of Periodic Bent DNA in Higher Eukaryotes.
Author	KIYAMA RYOICHI(Natl. Inst. of Bioscience and Human-Technol. Agency of Ind. Sci. and Technol.)
Source	Biophysics Journal Code:G0141A ISSN:0582-4052; CODEN:SEBUAL VOL. 40; NO. 3; PAGE. 173-178; (200006)
Figure&Table&Reference	FIG.4, REF.26
Pub. Country	Japan
Language	Japanese
Abstract	We found that a type of bent DNA appears periodically in the genomic DNA of higher eukaryotes and mapped their positions in a total of 93kb regions. Based on the structural characteristics and the relationship to nucleosome phases, this periodic bent DNA appears once in four nucleosomes on average and determines the phases first at these positions and then at the neighboring positions, thereby eventually aligning all nucleosomes on the genomic DNA. (author abst.)

図6 詳細画面

表3 JICST-EおよびJICST-E Plusのサービス別データ項目、検索タグ、料金および提供時間の比較

	J-East	JOIS		STN	
		コマンド	Enjoy JOIS	STN-International STN on the Web	STN Easy
データ項目 及び 検索タグ	整理番号/記事番号	△	○	○	○
	英文標題	○	○	○	○
	著者名	○	○	○	○
	著者所属機関名	○	○	○	○
	資料番号	△	△	○	○
	ISSN	○:注1	△:注1	○:注1	○:注1
	CODEN	○:注1	△:注1	○:注1	○:注1
	ISBN	○:注1	△:注1	○:注1	○:注1
	資料名	○	○	○	○
	巻・号	△	△	○	△
	ページ	△	△	△	△
	発行年	○	○	○	○
	資料種類	×	△	○	○
	記事区分	×	△	○	○
	言語	○	○:注2	○	○
	発行国	△	△	○	△
	英文抄録	△:注1, 注3	△:注1, 注3	△:注1, 注3	△:注1, 注3
	分類コード	×	△	○	△
	統制語	×	△:注3	○	△:注3
	フリーターム	×	△:注3	○	△:注3
料 金		無料	115円/分 125円/件	700円/回 150円/件	6,000円/時間 165円/件
提供時間		24時間	5時～翌2時:注4	24時間:注5	300円/回 291円/件

○:データ項目あり, 検索タグあり

△:データ項目あり, 検索タグなし

×:データ項目なし, 検索タグなし

注1:レコードによってデータがあるものとないものがある

注2:日本語と英語のみ検索可能

注3:自動切り出しおよびBasic Indexなどから検索可能

注4:月曜日 8時～翌2時(ただし, 月末は23時で終了) 日曜日 5時～19時

注5:日曜日 8時～22時は休止

```
#000001*    JST COPYRIGHT
CN  00A0589557
TI  Structure and Function of Periodic Bent DNA in Higher Eukaryotes.
AU  KIYAMA RYOICHI (Natl. Inst. of Bioscience and Human-Technol. Agency of Ind.
    Sci. and Technol.)
JN  G0141A (SEBUA) (0582-4052) Seibutsu Butsuri ( Biophysics )
VN  VOL.40,NO.3 PAGE.173-178 2000
CI  (A) (b2) (JA) (JPN) (FIG.4, REF.26)
AB  We found that a type of bent DNA appears periodically in the genomic DNA of
    higher eukaryotes and mapped their positions in a total of 93kb regions.
    Based on the structural characteristics and the relationship to nucleosome
    phases, this periodic bent DNA appears once in four nucleosomes on average
    and determines the phases first at these positions and then at the
    neighboring positions, thereby eventually aligning all nucleosomes on the
    genomic DNA. (author abst.)
CC  EC02020U (575.116.4)
KW  genome; repetitive sequence; nucleosome; globin; gene locus; DNA
FT  [bent DNA; beta globin; globin gene]
```

図7 JOISによるJICST-Eの出力画面

6. 従来のサービスとの相違点

J-EASTとJOISおよびSTN-Internationalとのデータ項目、検索タグ、料金および提供時間の比較を表3に示す。また、JOISによるJICST-Eの出力画面を図7に示す。

7. おわりに

J-EASTは平成12年12月からデータ更新など、運

用面の試運転を目的として、日本の業務時間帯を対象に、試験的提供を続けてきた。平成13年度からは、世界中に発信することを念頭に24時間運転を基本とした本格的提供を開始した。今後、J-EASTの利用が増えることにより、日本の科学技術活動が国際的に認知され、評価され、その結果として世界一流の人材や最新の情報がわが国に結集する状況を生み出すような、わが国の科学技術の基本である国際的な情報発信力の強化の一助となれば幸いである。