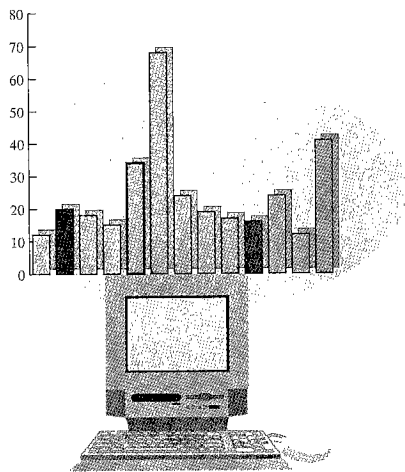


インターネット上のデータベース情報

横田一正

岡山県立大学情報工学部



● はじめに ●

インターネットの急速な発達によって、学術情報ははじめとするさまざまな情報源にネットワーク経由で容易にアクセスし利用できるようになってきた。電子メールからFTP、Webホームページへと、情報収集だけでなく情報発信の形態も変わってきた。このような環境は、個人的な情報の収集・発信にとどまらず、従来の研究スタイルを大きく変えつつある。

研究に有益な情報にはさまざまなものがあるが、よく利用したいのが文献情報である。インターネットを巡行しながら探すには以下のものがある。

- 文献アーカイブからキーワード等で検索する。
- 研究者個人のホームページを探す。
- 大学あるいは研究機関からたどる。
- 研究プロジェクトの成果を探す。
- 国際会議などの研究集会のホームページからたどる。
- 出版社あるいは雑誌のホームページから探す。

これで駄目なら、著者本人あるいは知っているような人にメールしたり、関連するメーリングリストに問い合わせる、という方法が考えられる。ACMのデジタルライブラリのように組織的に行われているものを除けば、上記のページにアクセスするまでは、さまざまな経験と試行錯誤が必要である。最近では研究集会の予稿集や雑誌のいくつかがオンライン化されており、単に2次情報だけでなく、1次情報も比較的簡単に入手できるようになってきた。たとえばデータベース分野で最大規模の国際会議であるVLDBでは、1996年以降会議録はオンラインで開催前から入手できるようになっている。

データベース分野では文献だけでなく、さまざまな情報が必要になったりする。

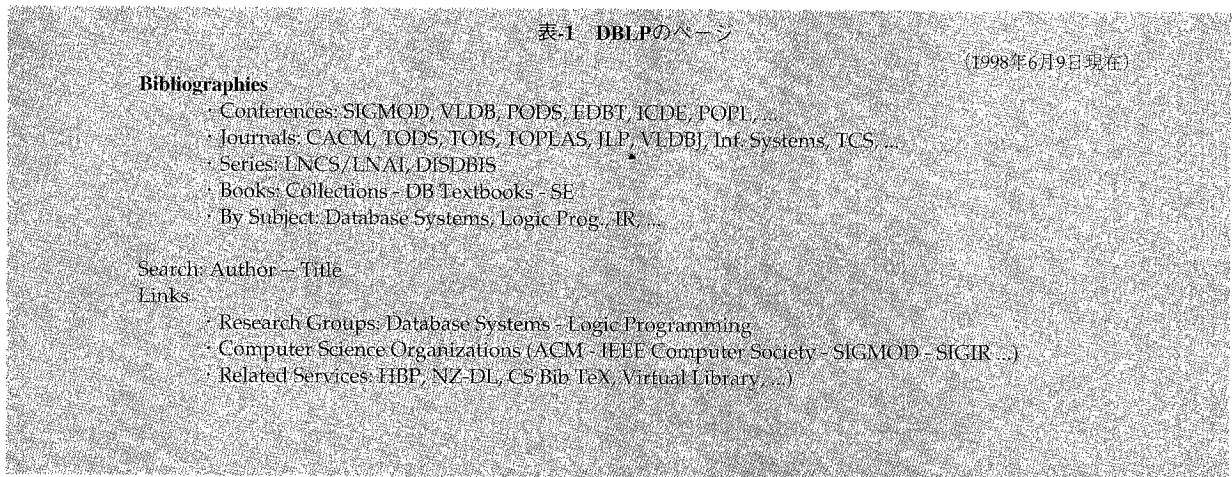
- 研究プロジェクト情報
- 人物や機関の情報
- システムやプロダクツの情報
- 研究集会の情報
- 出版情報
- 実システム例
- トラブル情報

これらの中で、プロダクツ、実システム、トラブル情報は公になりにくい性質を持っているので、入手するには困難が伴う。

本稿では、インターネット上のデータベース分野での情報源の主なものを紹介し、今後の日本での情報流通のあり方を考えるための材料を提供することを目的としている。ただしこのような情報源の内容は比較的頻繁に変更されるので、本稿が読者の目に触れる頃には異なったものになっているかもしれない。

表-1 DBLPのページ

(1998年6月9日現在)



いことをお断りしておく。ACMのデジタルライブラリなどのデジタル化の流れは別の機会に譲る。

● DBLP ●

ドイツのTrier大学のDBLP(computer science bibliography)

<http://www.informatik.uni-trier.de/~ley/db/>

は、データベースシステムと論理プログラミング分野での包括的なアーカイブを目指し、同大学計算機科学部のMichael Ley講師が個人で保守しているものである。“DBLP”はデータベースシステムと論理プログラミングの分野を示していると共に、Digital Bibliography & Library Projectの略ともなっている。ミラーサイトとして

<http://sunsite.informatik.rwth-aachen.de/dblp/db/>

<http://sunsite.ust.hk/dblp/db/>

<http://dblab.comeng.chungnam.ac.kr/~dolphin/db/>

がある。彼のホームページには“My main activity is to maintain the DBLP computer science bibliography.”と書かれているように、Ley講師は実に精力的に情報収集を行い更新しており、その活動は驚異的である。

ここには、データベースと論理プログラミング分野の約81,000の文献と約4,100のほかの研究者へのリンクがある(1998年4月現在)。さらに221の国際会議について、各会議ごとに、第1回から現在アナウンスされている最新のものまでの分かっているすべてのプログラム内容が記されている。雑誌については136種類のもの目次が載っているが、その中の約40については創刊号以来の目次がつけられており、残りのものには途中からの目次がつけられているかあるいはリンクが張ってある。継続出版物は5種類しかないが、その中の1つSpringer VerlagのLNCS/LNAIには最新のものまですべて記載されている。単行本についても

主要なものが収録されている。検索機能は、これら文献の著者と表題に対して行うことができる。

ホームページは表-1のように構成されている(1998年6月9日現在)。ただしこれは、原稿執筆中に変更になったもので、執筆開始時はもっと複雑なものであった。DBLPの説明(ドイツ語)は

<http://www.informatik.uni-trier.de/~ley/papers/cebit.ps>

にあるので興味ある方は参照されたい。

Michael Ley講師のこのホームページに対する努力に対して

ACM SIGMOD Service Award, 1997

THE VLDB Endowment 1997

Special Recognition Award

が与えられている。このホームページがデータベース分野でいかに大きな役割を果たしているかを示している。

情報は発信するところに集まるというのは、このDBLPでも示されている。国際研究集会を組織するとき、その広報活動は

- ホームページの作成
 - メーリングリストでの宣伝
 - DBLPなど主要アーカイブへの情報掲載
- が中心になりつつあり(このほかに、最後は個人的なコネに頼ることももちろんあるが)、DBLPへの掲載は必要条件といっても過言ではない。

ただし、筆者がチェックしたところ以下の問題も見つけた。

- 人工知能関係が少し弱い(たとえばICTAI, ECAI, FGCS'94など)。
- データベース関係でもいくつか洩れがある(たとえばADBIS'94, ADCなど)。
- 情報がdbworld等に流れたにもかかわらず、更新が大幅に遅れることがある。

これはやはり個人で保守を行うことの限界とも思われるが、このホームページが両分野の研究者にとって大いに有用であることには変わりはない。

表-2 主題別の論文数

	雑誌	会議	TR	合計
Others/Unclassifieds	127130	101416	12473	269368
Theory/Foundations of CS	64516	26237	3643	105494
Mathematics	54454	4016	4291	71650
Artificial Intelligence	21740	19884	6407	63521
Computer Graphics	27021	16368	1312	50838
Computer	18272	8921	4695	42321
Technical Reports	219	277	40664	41843
Parallel Processing	15087	11992	3896	35874
Databases	10035	11756	2090	28433
Distributed Systems	13450	4004	1887	28149
Softw.Eng./Formal Methods	10070	9842	2137	26234
Neural Networks	8877	6131	1438	20150
Operating Systems	7242	8848	492	18601
Human-Comp. Interaction	4018	7473	15	12453
Logic Programming	1955	4727	559	8336
Typesetting	3777	792	163	7149
Object-Oriented	1099	2601	822	5655
Wavelets	1307	365	353	2796
合計	392269	245650	87337	838865

表-3 データベース分野の分類

10519	DB systems
7663	DBs (query optimization, object-oriented)
2327	GIS Master Bibliography Project
1507	Information Systems
1307	entries relevant to DBMS
1099	entries relevant to temporal DBMS's
912	ACM SIGMOD Record, the newsletter of the ACM SIGMOD
514	ACM Transactions on DB Systems (TODS)
450	ACM SIGACT-SIGMOD Symposia on Principles of DB Systems
376	Bibliography relating to incomplete information in DB
327	the International Conferences on Very Large Data Bases
278	DB Group of the Univ. of Magdeburg
225	the Software Systems group at the Tech. Univ. Hamburg-Harburg
214	active DB
163	entries that simultaneously address space and time
147	ACM Transactions on Database Systems
89	the Int. Workshop on Advances in DB and Information Systems (ADBIS)
82	schema evolution (especially) in object-oriented DB systems
69	VLDDB 1994
58	DB research
50	entries that address spatial issue in DB management
45	views (especially) in object-oriented DB systems
12	The SYSLAB (Inf. Systems and Software Eng. Lab.) Bibliography Collection
28433	Total number of references

最近データベースとWebを統合しようとする研究が盛んに行われているが、このホームページが例題として取り上げられることが多くなっている。

● Karlsruhe大学 ●

ドイツの Karlsruhe大学の The Collection of Computer Science

<http://liinwww.ira.uka.de/bibliography/index.html>

は、同大学のPhD学生のAlf-Christian Achilles氏が保守している巨大文献アーカイブである。計算機科学

関係の文献(雑誌, 国際会議, TRがほとんど)が BibTeX形式で82万件以上収録されている(1998年4月現在)。その中で3万件以上は原論文のオンライン版にアクセスでき, 9千件以上は相互参照されている。

このアーカイブの特徴は, 文献がBibTeX形式で収録されていることと, 分類がしっかりしていることである。1998年6月9日現在の分野別の収録件数は表-2のようになっている。

各文献には, 著者, 表題, 出典などの基本的な情報のほかに, 抄録が約10%, キーワードが約30%, URLが約5%含まれている。上のDatabaseはさらに細かく分類されており, それぞれの収録件数は表-3の

表-4 アクセス件数

月	1998年6月9日現在			
	1995	1996	1997	1998
Jan		44645	78118	174538
Feb	2564	53232	92978	174201
Mar	31139	47281	102616	208484
Apr	22474	48564	113904	203889
May	22629	56526	103282	188435
Jun	27037	57504	93770	42460
Jul	29153	64172	105747	
Aug	28009	57644	79313	
Sep	29867	68195	129045	
Oct	42921	78471	166688	
Nov	37529	74044	161471	
Dec	25863	56710	140150	
月平均	27702	58895	113923	189909

表-5 ドメインごとのアクセス数

件数	割合	ドメイン
494281	14.69%	[未解決の数値アドレス]
484446	14.40%	.de (Germany)
424507	12.62%	.com (Commercial (mainly USA))
391896	11.65%	.edu (USA Educational)
195578	5.81%	.fr (France)
185095	5.50%	.uk (United Kingdom)
140314	4.17%	.net (Network)
105530	3.14%	.it (Italy)
84859	2.52%	.jp (Japan)
81183	2.41%	.ca (Canada)
66283	1.97%	.es (Spain)
62504	1.86%	.au (Australia)
50393	1.50%	.nl (Netherlands)
47294	1.41%	.se (Sweden)
39821	1.18%	.ch (Switzerland)
36556	1.09%	.kr (South Korea)
36317	1.08%	.fi (Finland)
36126	1.07%	.br (Brazil)
35887	1.01%	.dk (Denmark)
32324	0.96%	.at (Austria)

ようになっている。ただしこの分類は標準的なものではなく、意図は不明である。

これらの統計をしばらく続けて見ていると、分野によっては収録件数がかなり減少していることがある。分野が変更になったのか削除されたのか不明である。データベース分野を細かく見ると、特定のもの、たとえば1998年3月から4月にかけては上のPODS(the ACM SIGACT-SIGMOD Symposia on Principles of DB Systems)が200件ほど増えており、ほかはほとんど変化がない。PODSがオンライン化されたのともなって集中的に増えたことが予想される。

実際に検索で使用する時の問題は、DBLPと異なり収録範囲が明確でないことと、著者の名前(first name)の省略に一貫性がなく再現性に問題のある場合があることである。また、収録件数は多いが、DBLPの収録を必ずしもカバーしてはいない。TRの収録がこのホームページの特徴となっている。

またこのアーカイブの特徴は統計データが豊富にと

られていることである。ここには1995年2月以来の月ごとのアクセスの統計情報(1998年6月9日現在)があるので表-4に添付しよう。

ドメイン別のアクセス数も見てみよう。(不明も独立のものとしてカウントすると)全部で141のドメインごとのアクセスの統計がとられている。アクセス数の少ないドメインに想像を働かせているのも実に楽しいが、紙面の都合で上位20位(全体の90%)までを表-5に載せよう。興味ある方は実際にアクセスしてみたい。

● ACM SIGMOD ●

米国のサイトを1つ紹介しよう。データベース分野でもっとも充実しているのがACM SIGMODのホームページ

<http://bunny.cs.uiuc.edu/>または

<http://www.acm.org/sigmod/>

であり、もっとも包括的な情報を持っているといえる。SIGMOD自身の情報のほかに多くの情報を持っている。

国際会議の一覧表が

<http://bunny.cs.uiuc.edu/sigmod/conferences/>

に1993年以降カレンダー順に並べられていた。アドレスを見て分かる通り、このホームページを管理しているのはイリノイ大学であった。参考までに1998年3月現在掲載されていたそれまで1年の会議数は表-6(a)の通りであった。これは、一目瞭然に多くの国際会議を見ることができ、投稿先・参加先を検討するのに便利がよかった。DBLPが会議別に掲載されているのとは対照的といえよう。しかし、1998年の掲載会議数が少ないと思っていたら、1998年4月からはSIGMOD独自のページを廃止し、

<http://www.cs.auc.dk/~busatto/db-events/main.html>

にリンクを張るだけとなった。このホームページは論文募集と参加募集を分離し、前者は締切ごとに、後者は開催日ごとに並べ替えられた。1998年4月以降開催される会議で掲載されているものの数は表-6(b)の通りである。データベース関連の国際会議の数の多さに驚かされる。このホームページはデンマークのAnaborg 大学計算機科学部のRenato Busatto助手が保守しているもので、彼はほかに時制データベースのアーカイブも管理している。後述するメーリングリストdbworldも独自にアーカイブを持っており、それから抽出された会議情報が別の人によって保守されていたが、それも1997年にこのホームページに吸収されている。ACM SIGMODほどの組織と財政の基盤を持っていても、このような情報の保守は個人のホームページに依存せざるをえなくなったことは、ネット

ワーク上の情報源と考える上で実に象徴的である。

ACM SIGMODのホームページには、そのほかには

- パブリックに利用可能な46個のデータベースシステムの一覧表とリンク
- 26のデータベースの研究グループあるいはプロジェクトへのリンク
- データベースパブリケーションサーバへの約30のリンク

などが掲載されている。以前はデータベース分野の研究ファンドの一覧表も掲載されていたが現在は消えている。

ACM関連の国際会議や雑誌はデジタルライブラリとして1次文献にアクセスできるようになったが、SIGMOD RECORDなど一部を除けば会員のみ利用可能となっている。データベース分野で最大の国際会議VLDBが1996年以来会議の予稿集を完全にオンライン化したのに比べると対照的である。ACM以外の文献については上記ドイツのサイトに比べるとないに等しい。しかし他のアーカイブなどへのリンクなど包括的である点で利用価値が大きい。

日本では比較的最近、情報処理学会データベースシステム研究会のホームページが神戸大学

<http://www.db.cs.kobe-u.ac.jp/IPSJ-DBS/>

で、電子情報通信学会データ工学委員会のホームページが東京大学

<http://www.tkl.iis.u-tokyo.ac.jp/OtherInfo/DE/>
で運用されており、今後の充実を期待したい。

● メーリングリスト ●

米国ウィスコンシン大学のRaghu Ramakrishnan教授が運営している世界的なメーリングリストdb-worldは参加者が3,000人を超え、論文募集、参加募集、人材募集などを中心としている。データベースの特定の問題に関する質問のように一部の人しか興味を持たないものは原則として禁止している。投稿数は平均すると月200通くらいである。

参加申込み:majordomo@cs.wisc.edu

アーカイブ:<http://www.findmail.com/listsaver/dbworld/>

一方、国内向けのデータベース分野でのメーリングリストとして、筆者は、1990年9月からdbjapanを運営している¹⁾。

参加申込み:dbjapan-request@c.oka-pu.ac.jp

参加数は約900人で、投稿記事には特に強い制限は設けていないが、月平均20数件の投稿がある。登録者の名簿管理をやっているため、そのアーカイブ

<http://alpha.c.oka-pu.ac.jp/dbjapan/>

は登録者のみアクセスできるようにしている。それに

表-6 国際会議一覧表

(a) SIGMOD

年	1997								1998							
月	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	
件数	8	17	16	15	15	15	10	8	2	5	6	4	3	7	6	

(B) Busatto氏

月	3	4	5	6	7	8	9	10	11	12
件数	11	16	16	32	23	38	22	20	12	6

関連したページ

<http://alpha.c.oka-pu.ac.jp/~yokota/db.html>

にはデータベース研究に関連するさまざまな情報源のリンク集を置いており、自由にアクセスできるようになっている。dbjapanの参加者は最近では微増状態、ほぼ飽和しつつあると考えている。ACM SIGMOD日本支部のメーリングリスト等と突き合わせると、ネットワークを使ってアクティブに活動しているデータベース研究者や利用者の数は、日本には約1,500人いると考えている。

● おわりに ●

本稿で見たように、さまざまな情報源のほとんどは、多くの個人の個人的な情熱と努力という不安定な基盤の上に成り立っている。最近学会のホームページが充実してきたが、ACM SIGMODのところで述べたように組織外の情報については貧困といわざるを得ない。さらに、組織で保守していた情報も個人のホームページへのリンクに切り替えつつある。これまでのところ、個々の組織を超えた包括的な情報を維持するためには、不安定ながらも個人の頑張りに依存せざるを得ないのが現状である。筆者の運営しているdbjapanも組織的な運営に切り替えようと計画したこともあるが、最終的には個人に依存せざるを得ず、山積する問題点を確認しただけに終わった。今後、現在は充実しているが本質的には不安定な情報源をいかにサポートしていくかが大きな問題といえるだろう。

現在データベース分野で、Webを含む異種分散情報源を1つの仮想的なデータベースシステムとするための研究が盛んに行われている。Webの広がりデータベース研究とは独立に行われてきたということ、また最近でもXMLの提案にデータベース研究者が関与していないこと、など、データベース研究に携わるものとして大いに反省すべき点があると共に、今後の研究方向に大きな示唆を与えてくれる。

参考文献

- 1) 横田一正, 谷 園子: 電子メーリングリストdbjapan の発展経過と今後の展望, データベース, 関西データベース協議会(1995).
(平成10年5月6日受付)