

ソフトハウスが始めた手作りデータベース —HiNET&HAPPINESS 開発秘話—

杳澤 尚明*1, 飯田 一幸*2

1982年5月に専門紙の二次情報を提供する手作りのテクノサーチで開始した商用データベースHiNETは、日本語処理システムHAPPINESSのキーワード自動抽出をツールに、雑誌論文のタイトルサーチや各種新聞記事の全文提供、マーケティング情報や音楽CDなどにメニューを広げた。HiNETは操作性の追究やコード変換のトラブル、ネットワーク構築、他サービスとの連携など大小様々な課題解決を経てG-Searchに引き継がれる。またHAPPINESSは独立した製品販売事業として図書館市場に展開するほか、検索機能を装備したシステム構築ツールに拡大し、多くのユーザを獲得した。

キーワード：平和情報センター, HiNET, HAPPINESS, G-Search, 商用データベース, 日本語処理システム, キーワード自動抽出, テクノサーチ, 新聞記事, ネットワーク

1. はじめに

公的機関やメーカーではない一介のソフトハウスである(株)平和情報センター(平和生命保険(株)の電算部門が1972年2月に独立。略称HIC)が始めた商用データベースHiNET(HIC Information Network)と、それを支えた日本語処理システムHAPPINESS(HIC Advanced Program Package for Intelligent Nihongo Engineering by Syntax and Semantics)について、システム担当者の立場から当時を振り返ってみる。HiNETは1991年1月に(株)ジー・サーチのG-Searchサービスに引き継がれ現在も稼働している。またHAPPINESSはプラットフォームを超えて現在でも販売を継続している超ロングセラー製品である(表1)。

2. HiNET 事業の立ち上げ

商用データベースサービスの草創期(1970~1980年代)においては、オンラインで入手できる情報は極めて限られていた。そのため、企業の調査部門では、新聞記事等を自ら保管しておくか、図書館等の公共機関を利用するか、さもなくば情報の保有者に直接アクセスするなど、大変な苦勞があった。

1979年に、新聞記事を情報源にリサーチ業務を行っていた竹谷一郎氏(インフォマティックリサーチ社長, 故人)より、マイクロフィッシュに記録された新聞情報を簡単に

引き出す方法はないかとの相談を受けた。その検討過程で、これは一企業の社内システムに留めず、公開することが大きなビジネスチャンスになると捉え、福島芳直氏(故人)が中心となってHiNET事業はスタートした。

HiNETは、将来の総合データベースサービスを目指して、まず、新聞記事データベースの充実を図ることから着手した。

2.1 情報(データベース)を商品に

HiNETの立ち上げ時(1980年代前半)は電子化された情報が少なく、電子化されていても印刷を目的とした専用フォーマットであったり、最初の入力データはあっても、その後の校正が反映されたものではないなど、流用が困難であることが多かった。さらに情報所有者の多くは「オンラインでコンテンツが提供されると印刷物の販売に大きく影響してしまう」という意識に支配されていた。

このような環境の中で、HiNETでは企業の研究開発部門に向けた新聞記事データベースの提供を企画し、国内の主要な産業・工業新聞5紙をデータベースの対象として開発に着手した。電子化されたデータがないため、データの作成から手がけることになった。従ってフルテキストなど望むべくもなく、記事タイトルと抄録からなる二次情報の提供であった。これが1982年5月にサービスを開始した「テクノサーチ」である。情報としては脆弱であっても新聞記事をオンラインで探すことができる(検索結果を元に、一次情報は縮刷版等で得られる)意味は大きく、多くの企業に受け入れられた。

テクノサーチにおいては、公衆漢字端末機標準仕様に準拠する端末を用い、公衆回線を通してHiNETセンターに直接接続する形態であった。

料金体系としては、検索料金(50円/0.2ユニット)と出力料金(70円/件)からなる従量課金制であった。ここでユニットとは、検索処理によりコンピュータ資源(CPU,

*1 かつざわ なおあき (株)平和情報センター
〒112-0002 東京都文京区小石川 1-3-21 日本生命春日町第2ビル
Tel.03-5802-2648

*2 いいだ かずゆき (株)ジー・サーチ
〒108-0022 東京都港区海岸 3-9-15 LOOP-Xビル
Tel. 03-5442-4735 (原稿受領 2008.7.8)

I/O装置など)をどのくらい使ったかを示す単位である。

入力は英字・カナに限定し、検索結果の表示は漢字に対応した。当初は、端末への出力のみであったが、1987年2月にオフライン出力サービスを追加した。

検索用に、技術、経営・社会、機関名の3種類のキーワードと分類コードを手作業で付与しており、これを検索に供した。つまり統制語方式であるが、抄録作成に加えてキーワード付与作業や用語辞書の整備は、非常に手間のかかるものであり、新しいデータベースをメニューに追加する上でのネックであった。

そこでこの解消と利用者の立場を考慮し、自由語方式に切り替えることにした。そして、そのインデクシング作業の自動化を図るために開発したソフトウェアがHAPPINESSである。

HAPPINESSの適用により、大幅なコストダウンとタイムラグ短縮の見通しが立ち、HiNET事業の拡大に乗り出すことができた。1983年7月にサービスを開始した「タイトルサーチ」(科学技術系雑誌の論文タイトルを収録したデータベース)はその第1号である。

2.2 コンテンツの収集

待ちに待った新聞<フルテキスト>データベースの提供は1986年4月に開始した「朝日新聞記事データベース」である。本提供開始に先立ち、1ヶ月間の無料モニターサービスでは非常に大きな反響を呼び、問い合わせの電話が鳴り続く状況であった。記事データは朝日新聞社よりオンラインで提供された。またキーワードはHAPPINESSによる自動付与であるため、人手を介さずタイムラグの非常に短いサービスを実現することができた。

朝日新聞のサービスを機に、各プロデューサの動きも活発化し、1987年以降、日経産業新聞、読売新聞などの新聞記事や、マーケティング資料、中国情報、知的所有権判例、音楽CDなどデータベースの種類は急速に拡大していった(表1)。

3. HAPPINESS 事業

前述のように、HAPPINESSはHiNETを支援する運用ツールの役割を担ったが、本来は販売を目的としたソフトウェア製品である。そのため体制もHiNETとは独立した事業形態を取った。コンピュータで日本語が扱える時代を迎えた新たなビジネスチャンスに、一方は情報提供サービス事業としてHiNETを、もう一方でソフトウェア製品販売事業としてHAPPINESSを立ち上げたわけである。

HAPPINESSの狙いは、拡大するであろう日本語情報処理に有効となる基本機能の提供であった。つまり単にキーワードを抽出するだけでなく、様々な分野に適用できる製品を目指した。その主要機能が分かち書き(形態素解析)とフリガナ付けである。これは開始時点から今日に至るまで変わっていない。逆に言えば当時としては極めて先進的な試みであった。

3.1 日本語との格闘

HAPPINESSは1981年6月より開発に着手したが、トライアル&エラーの連続であった。ともかく日本語を知らねばならない。しかし当時のコンピュータ関連の書籍には日本語を扱っているものが殆どなく、わずかに写植の話が散見される程度。外国文献にいたっては日本語を相手にしていない。結局、国語辞典をバイブルとし、コンピュータとは無関係に日本語の文法に関する書籍を読み漁って、独自の解析手法を蓄積していった。また処理ロジックに外れた例外を「辞書」に登録するという地道な作業が限界に達したところ、国語辞書データの入手が可能であることを知り、全面的な再構築を決断するなど、2年足らずの間に3度の書き直しを実施した。

こうして1983年4月からHiNETのタイトルサーチでのトライアルを開始。辞書の拡充やロジックの改訂を繰り返して、検索結果の妥当性を検証するなどして、ようやく実用にこぎ着けた。そしてタイトルサーチのサービス開始を見届けた同年9月、製品販売事業のスタートを切った。

3.2 市場の拡大

当初は放送局の映像ライブラリシステムや財務会計情報の提供サービスなど、やはりデータベース向け需要に適用されたが、1984年5月の国立国会図書館への提供を皮切りに、図書館市場へ大きく展開することになる。それは図書館が作成する書籍の目録に「アクセスポイント」と呼ばれる書名のカナ分かち書きが必要であり、それにHAPPINESSの処理結果がそのまま使用できたからである。国立国会図書館にはこれ以降、プラットフォームを超えて現在でも利用いただいている。

さらにソフトウェア販売事業は、HAPPINESSを中核とした製品拡大を図った。HAPPINESSの販売促進用デモンストラシステムを発展させ、データベースの定義と入力画面や検索画面のイメージを定義すれば、直ちに検索機能を持つアプリケーションシステムが使用できるという、統合情報管理システムBride/Happiness (BRoad Information manager Designable Equipment with HAPPINESS)である(1987年4月リリース)。当製品で構築したアプリケーションは、入力したデータを登録すると、自動的にHAPPINESSでキーワード抽出が行われ、直ちに検索対象となる。つまり特別な検索技術がなくても検索システムの構築ができるため、企業の研究開発部門や情報管理部門を中心に多くのユーザを獲得することができた。当時の大型汎用機向け製品としては画期的であり、1990年にはSOFTIC (Software Information Center)よりソフトウェア・プロダクト・オブ・ザ・イヤーで表彰されることとなった。

4. HiNET のシステム

この章では、HiNETを特徴付けるネットワーク構成を中心に、オンラインサービス環境の進展について説明する。図1は1987年当時のシステム概要である。

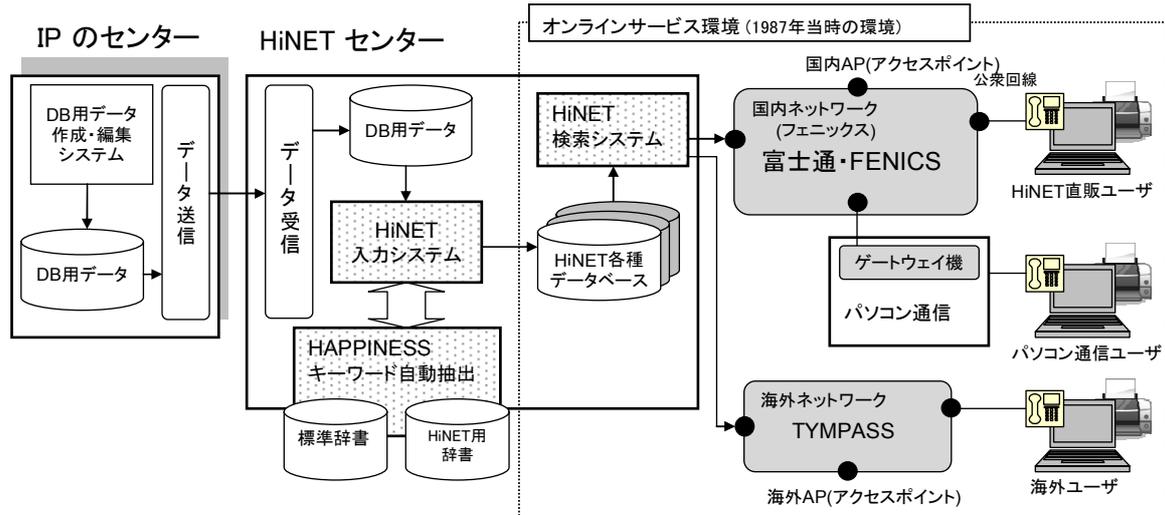


図 1 HiNET システムの概要図

4.1 検索システム

1982年に公開した検索システムは、当初サーチャーを想定したコマンド方式を採用した。当時の情報検索は、適切なキーワードの決定など専門的な知識を持つサーチャーの仕事であった。しかし朝日新聞記事データベース（1986年）以降は、一般ユーザの利用を想定して、システム誘導型のメニュー方式にも対応した。

(例)

次の処理を選択して下さい。

1. 思いつく言葉で検索する。
2. タイトル中の言葉で検索する。
3. 日付（年，年月，年月日）で検索する。
- P. 検索結果を出力する。
- H. 補助機能を利用する。

->

検索システムの命題は操作性の良さであるが、当時の英字カナしか入力できなかったテキスト端末では、さらにシンプルな形態が要求された。例えば、日付検索のカスケード処理（範囲指定を含め年・年月・年月日の組み合わせに対応）、コマンドの簡略化、ローマ字カナ変換、入力値の全角・半角同一視などである。また検索結果の表示には禁則処理を組み込んで読みやすくするとともに、全角・半角の切り替えを極力抑えることにより、端末の印字速度を落とさない等の対応を行った。

公衆通信漢字端末機の日本語入出力には、予想外の苦労があった。ホストコンピュータの日本語処理は富士通のJEFコード、端末機はJISコードと異なるが、ビット操作で変換できると考えた。しかし実際に端末を接続してみると、検索結果は意味不明な文字の羅列でしかなかった。

検索システムではJEFからJISに変換しているが、ホストコンピュータの通信制御装置で稼動するネットワークコントロールプログラム（NCP）が、データを回線に送り出す際にコード変換を行っていたからである。NCPは端末

と通信するための各種条件を管理しているが、公衆通信漢字端末機を日本語端末として認識せず、単なるASCII端末として扱うためであった。つまりデータは全て1バイトであり、文字種を判断しない。さらに表示文字や制御用コードが割り当てられない領域はNULLコードに変換されてしまう。

この対応として、検索システムの中では、NCPを通して変換されたコードがJIS漢字コードになるような、仮想的漢字コード体系をつくり、そのコード体系に変換するよう修正を行った。またNULLとなる領域には自らコードを割り当て、変換テーブルを修正した（図2）。

4.2 検索用端末

データベース普及のネックは端末環境であった。HiNETを利用するためには、1980年に発表された公衆漢字端末機標準仕様に準拠する端末が必要であった。仕様が発表されて間もないために、あまり普及しておらず、データベース契約とともに、端末購入が必要であった。「HiNETの便利さは理解できたが、利用する端末がない。100万円近くする専用設備をすぐに用意することはできない」という言葉を何度も耳にした。

しかし、その端末環境も1986年ころから始まったパソコン通信の急激な普及が大きく影響し、多様化していった。パソコン通信ソフトの主なものとしては、G-POT（紀伊國屋書店）、DCOM（電通国際情報サービス）、PCOM（パーソナルビジネスアシスト）、まいとーく（インターコム）などがあった。またこの頃「オートパイロット」という言葉がよく使われたが、これは、サービスへのログイン（接続）、検索、結果の出力、サービスのログオフ（終了）など一連の処理を自動で行ってしまうという通信ソフトの機能である。当時の低通信速度では、非常に便利な機能であった。

4.3 通信ネットワーク

HiNETの開始時は公衆通信回線（電話回線）で直接接

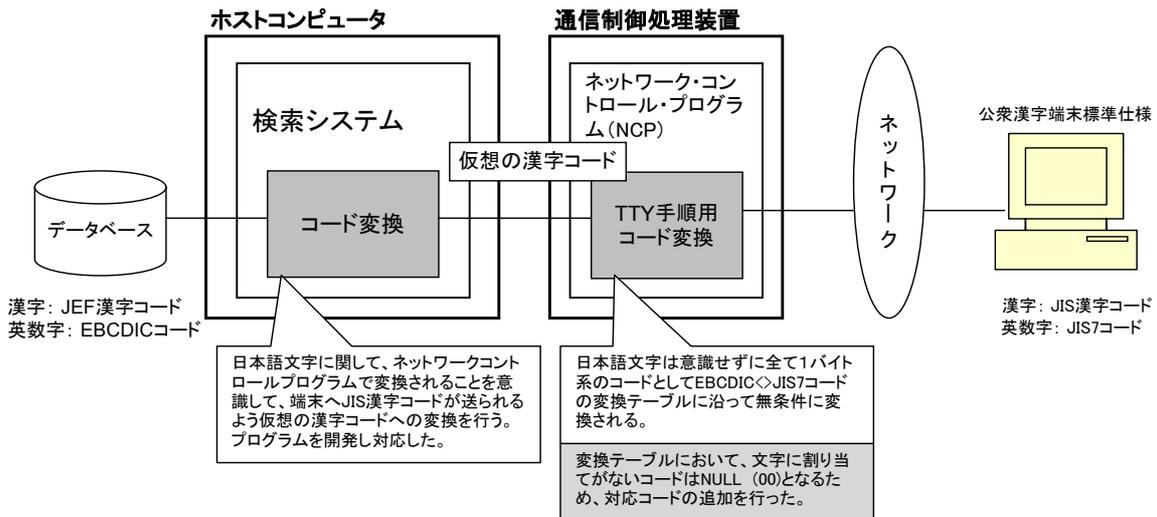


図2 日本語コード変換手順 (1981年対応)

続していただいた。ソフトハウスには専用のネットワークもなく、全国どの地域からでも東京にある HiNET センターにつないでいただく形態であった。会員数が増える中で、利用者からは、せめて大都市にアクセスポイントを用意して欲しいとの要望が寄せられていた。

1987年7月に回線自由化が始まり、メーカーほか通信各社が VAN サービス事業を立ち上げ、公衆-特定-公衆の接続を用いたネットワークが急速に広まっていった。HiNET では、富士通の VAN サービスである FENICS (FUJITSU ENhanced Information and Communication Services) と接続を行い、全国ネットワークを提供した。海外に対しても TYMPASS との接続を行い対応した。通信速度も目で追える、読める程度の 300bps から 1,200bps, 2,400bps と、高速化が図られた。

4.4 パソコン通信との連携

VAN サービスと連動したパソコン通信サービスは、データベースサービスと同じネットワーク環境、同じ端末環境で利用できるものであり、親和性の高いサービスであった。そこで HiNET は 1987年7月、同じ FENICS をネットワークとして利用する NIFTY-Serve とゲートウェイ接続を行い、NIFTY-Serve の会員であれば、HiNET と契約しなくてもデータベースの利用を可能とした。パソコン通信の利用者は主に個人会員であり、HiNET から見ると新規ユーザ層の開拓、NIFTY-Serve から見ると、サービスの充実というコラボレーションであった。

ここで養われたゲートウェイ技術が活かされ、後に HiNET を引き継ぐ G-Search サービスの特徴となった。

5. 課題と運用に迫られる日々

HiNET 事業は課題解決の歴史であった。開発面では、

- ・テクノサーチのタイムラグ短縮 (1985年)
- ・オフラインプリント機能の提供 (1987年)
- ・新聞横断検索機能の提供 (1988年)

- ・一次資料のオーダ受付機能 (1988年)
 - ・サービス時間の拡大 (1990年)
- など。またデータベースの運用では、
- ・毎日送られてくるデータの確認
 - ・データベースへの登録・確認
 - ・オフラインプリントの発送
 - ・キーワードの辞書管理と用語集作成
 - ・HAPPINESS ユーザ辞書の整備
 - ・毎月の請求書の作成
 - ・トラブル発生時の緊急対応
- など、多くの作業に日々追われた。

開発作業・運用作業をわずか数名で対応するため、まさに休む間もないといった状況であった。

また、コンピュータ資源の確保も大きな課題であった。今では、パソコンでさえ数百 GB のディスクが搭載されているが、当時 GB 単位のディスク装置を必要とするデータベースは、ある意味大規模な装置産業であった。

6. データベースのこれから

インターネットの普及により、ほとんどのデータベースサービスがインターネットを基盤としたサービスに変わっており、高速なネットワーク環境のもと、提供される情報も、テキスト・画像・動画など多様化している。

また企業の活動を見ると、社内の情報共有やコミュニティを目的としたポータルサイトの構築が進んでおり、ここでもデータベースは大事な役割を担っている。

別の角度から見ると、企業のプロモーション活動においても「検索」が重要なキーワードとなっている。Google など主要な検索サイトの中で、自社サイトの露出をいかに高めるか、SEO (Search Engine Optimization) 対策などの取り組みが行われている。本来、情報を管理し、正確に検索させるのがデータベースの使命であるのに、検索対象であるはずの情報が、データベースに合わせて編集・加工されるといった逆転現象まで生じている。

時代は変わっても、情報を管理して、利用者に必要な情報を提供するといった、データベースの目的は変わらない。しかし、扱う情報はさらに多様化し、利用者層の拡がりに応じて、ニーズも拡大するものと考えられる。氾濫する情報の中から、効率よく、本当に必要な情報を探し出すといった基本的な機能に加えて、検索で得られた情報を、データマイニング技術で分析するなど、付加的な機能も求められてくる。今後も様々なニーズや課題を解決しながら、さらに進化したサービスが実現されるものと期待する。

7. おわりに

今回の寄稿にあたっては、HiNETとHAPPINESSの2つの事業の関わりを中心に記述させていただいた。筆者2名は事業開始の決定を受けて参画したメンバーであり、決定に至るまでの議論や検討過程を掘り起こすことができなかったことは残念である。また事業の立ち上げには多くの方が参加されているが、紙面の関係で紹介しきれなかったことをお詫びしたい。2つの事業とも形こそ変われ四半世紀以上に渡って続いている。改めて故人となられた竹谷氏や福島氏をはじめ、当時のメンバーの方々の功績に敬意を表したい。

■ 用語解説：日本語テキストの自動分かち書き ■

欧米語と異なり、日本語ではテキスト中の単語の切れ目が明示されていない。日本語テキストから意味のある単語を索引語として抽出するには、人手または機械による分かち書きが必要である。そのための自動分かち書きソフトウェアのひとつがHAPPINESSであ

る。そのアルゴリズムは一般的に形態素解析と呼ばれるものであり、Google、Yahoo!などの検索エンジンでも形態素解析が用いられている。簡単に利用できる形態素解析ソフトウェアとしてはChaSen(茶釜)などがある。

Series: Footsteps of information retrieval service pioneers(6): The handicraft database which software house began - The HiNET & the HAPPINESS development secret story. Naoaki KUTSUZAWA (Heiwa Information Center CO., LTD. Nihonseimei Kasuga-cho Dai2 BLDG. 1-3-21 Koishikawa, Bunkyo-ku, Tokyo 112-0022 JAPAN), Kazuyuki IIDA (G-Search limited. Loop-X Building, 3-9-15 Kaigan, Minato-ku, Tokyo 108-0022 JAPAN)

Abstract: Commercial database HiNET that began on a handicraft TECHNO-SEARCH, in May 1982, that offered the secondary information on the trade papers, at a late year that used keywords automatic extraction of Japanese processing system HAPPINESS in a tool, expanded the menus -- TITLE-SEARCH (the magazines treatises), various newspaper articles (full text), marketing information, and music CD, etc. --. HiNET settled the big problems and the small problems, such as investigation of the operability, the trouble of the code conversion, network construction, cooperation with the other service, and it was succeeded in G-Search. On the other hand, HAPPINESS unfolded as an independent product sale business, for instance, in the library market, and later enlarged in a systems construction tool equipped with search feature, and acquired many users.

Keywords: HIC / HiNET / HAPPINESS / G-Search / commercial database / Japanese processing system / keywords automatic extraction / TECHNO-SEARCH / newspaper articles / network

表 1 HiNET と HAPPINESS の歩み

		HiNETサービスの出来事	HAPPINESSの出来事
創	1981年 12月	HiNET仮提供開始 工業、産業専門5紙の記事情報『テクノサーチ』を公開	
	1982年 5月	HiNET本提供開始	
	1983年 7月	科学技術系雑誌の論文タイトル情報『タイトルサーチ』を追加	
生	9月		日本語キーワード抽出システム『HAPPINESS』販売開始
	1984年 4月	HiNETの販売で紀伊國屋書店と提携	
大	1985年 5月	日本データベース開発を設立。テクノサーチのインデクシングを対応	
	10月	富士通のVANサービス「FENICS」に接続 アクセスポイントの拡充と品質向上	
	1986年 4月	『朝日新聞記事データベース』を追加	
	5月		第1回「HAPPINESSユーザ研修会」を開催
	1987年 2月	オフライン出力サービスを追加 (テクノサーチ、タイトルサーチ、朝日新聞、日経産業新聞)	
	3月	マーケティング資料の所在情報データベース 『マーケットサーチ』を追加	
	4月	国別経済貿易情報『ジェット・エース』を追加	データベース構築支援システムパッケージ『Bride』販売開始
	7月	パソコン通信「NIFTY-Serve」とゲートウェイ接続	
	9月	『読売新聞記事データベース(YOMIDAS)』を追加	
	10月	『朝日新聞夕刊オンライン版(速報)』を追加	
飛	1988年 4月	『日経新聞記事データベース』を追加	
	11月	『毎日新聞ニュースリリース』、雑誌『AERA』を追加 早稲田大学にホスト間接続を行いHiNETの提供を開始	
	1989年 1月	海外データベース『INFOCUE』を追加 NIFTY-Serveからゲートウェイ提供(双方向G/W実現)	
	3月	中国の新聞記事情報『CDC中国情報』を追加	
	4月	『毎日新聞記事データベース』を追加	
躍	7月	HiNET統合データベースサービス『G-Search』提供開始 『中国情報速報サービス』を追加	
	9月	知的所有権判例データベース『インテレクト』を追加	
	10月		WS向け日本語文書情報系データベースシステム パッケージ『Future/Happiness』販売開始
	1990年 4月	23時間サービスを開始、サービス時間を大幅に延長	
	9月	音楽出版社の『音楽CD総カタログデータベース』を追加	
	11月		『Bride/Happiness』 “ソフトウェア・プロダクト・オブ・イヤー'90”を受賞
	1991年 1月	(株)ジー・サーチ(商用データベース専門会社)を設立、『HiNET』を移管	
4月		統合情報管理支援システムパッケージ『SUPER Bride』販売開始	
10月		『HAPPINESS』100本販売達成	