

東京スポーツ新聞社

COMPANY PROFILE

会社名	東京スポーツ新聞社
創刊	1960年4月1日
本社	東京都江東区越中島2-1-30
代表	代表取締役社長 太刀川 恒夫
発刊紙	発刊紙 東京スポーツ、中京スポーツ、大阪スポーツ、九州スポーツ
発行部数	2,424,600部（4紙合計、2005年11月30日現在）
事業	東京本社を中心に、5カ所の拠点印刷網で全国を夕刊3紙、朝刊1紙でカバーする全国紙を手掛ける。1960年の創刊以来、「ライバルは駅売店などで併売されるお茶やお菓子」というユニークなスタンスで、一貫して宅配よりも即売に主眼をおいた販売体制をとる。また「スポーツ娛樂紙」を標榜し、一面の強烈な見出しを筆頭に独自の切り口による紙面づくりで、数多くの読者から支持と親近感を得ていることでも知られている。
URL	http://www.tokyo-sports.co.jp/



情報システム部
部長
吉良 輝夫 様

Category

業種ソリューション

Solution

新聞トータルソリューション

Product

TX1

Innovation Report

コンテンツデータベースシステムのその先 高速で大容量の新世代XMLデータベースが 「東スポ」をますます面白くする

今日の出来事を、今日の紙面に活かす。
東芝ソリューションのXMLデータベース「TX1」をエンジンとした
コンテンツDBシステムが、「東スポ」をもっともっと面白くする。

5年分の全記事を

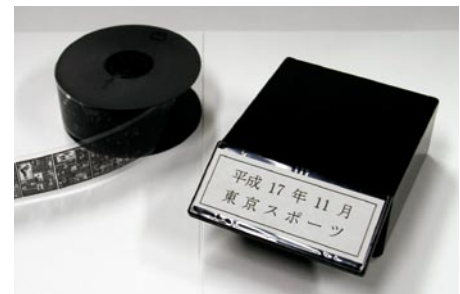
秒速 **150** 万件で

導入背景

XMLデータベースのいま

XMLデータベースとは

近年のIT化でもっとも変わった点のひとつは情報のデジタル化だ。かつて文書や伝票の形で流れていた情報はデジタルデータに取って代わりつつある。それに伴い、企業が扱うデータの量は飛躍的に増加している。そのようなデジタルデータを保管、検索する集合体であるデータベースと無縁の企業は、皆無と言っても過言ではない。一方データベースの設計には検討すべき点が多い。データベースは、設計時にデータ構造を予め定義しておかなくてはならない。しかし、一度決めたデータ構造を、仕様変更のたびに加工することは、全体の設計はいかに及ばず、アプリケーションにまで影響することも考慮しなければならない。例えば、製品コードの桁数を変更する、という一見簡単そうな変更にも、かなり綿密な作業が必要になる場合がある。しかも、企業が扱う多くの情報はたえず変化する。数値や項目数の決まった定型データだけでなく、文書や画像が混在する非定型データの取り扱いも無視できない。これらのデータ構造の違いを考慮せず、過剰にデータ設計をし過ぎると、結果的に全体の負荷が大きくなってしまふ。文章や画像が混在する非定型データであればなおさらだ。データベースの構造設計はノウハウの塊、と言われるゆえんだ。これまでは、必要に応じて専用データベースを設けたり、データ構造を変更して対応してきた。しかし、データベースの加工や変更には手間がかかるうえ、アプリケーションの処理時間への遅延にもつながりかねない。その解決策として、数値系データなどの定型データは、リレーショナルデータベースで管理し、文書や画像などの非定型データは、マークアップ（印つけ）言語のXMLデータを使ったXMLデータベースで管理する最適化が重要なポイントとなっている。



過去の新聞を記録したマイクロフィルム

将来展望を視野に入れXML化を促進

非定型データの取り扱いにXMLデータベースが注目を集める理由は、その柔軟性にある。XMLデータ形式であれば、一般的なデータベースのように項目ごとに最大バイト数まで厳密に定義された形式に当てはめることなく、データをそのまま格納し、利用できる。ま

た、使っている最中にデータの構造を変えることも比較的容易だ。これらの特徴が注目され、XMLデータベースの利用が広がりがつつある。膨大な情報を扱う報道機関では、その動きが活発になっている。ニュース素材は、写真、記事（テキスト）、音声、動画などニュースによってまちまちで、融通の利くXML形式のファイルが、こうしたニュース素材の配信や管理に適しているからだ。ロイター通信社がインターネットでテキスト、画像、動画などのニュース素材を配信する方法として、XMLをベースにしたNewsMLを採用し、普及に熱心だ。その後、日本でも徐々に普及しはじめ、大手各紙が新幹線の電光ニュースの送信形式に採用したり、2003年に共同通信がNewsMLでの配信を開始している。独特な切り口でスポーツや芸能、社会娯楽などを報道する紙面が多く読者に愛され、関東以北では「東スポ」（中部、東海地方では「中京スポ」、関西では「大スポ」、九州では「九スポ」）の愛称で親しまれている、東京スポーツ新聞（<http://www.tokyosports.co.jp/>）もXMLデータベースで紙面データベース構築に乗り出した企業のひとつだ。

答えは東スポの面白い紙面にあり

トップニュースでは分からない 人々の暮らしを語るデータベースを

東京スポーツ新聞社がXMLデータベースの構築を検討しはじめたのは、2004年11月のことだった。「インターネット時代になり、紙面データベースを一から作る必要を感じていました。」と紙面データベース構築の背景を話すのは、同社の情報システム部部長の吉良輝夫氏だ。同社は印刷を外部に委託しており、新聞紙面制作システムもその印刷会社のものを使っている。紙面制作に必要な記事データベース、画像データベースは、すでに開発済みだった。しかし、新たに構築するデータベースは、これらの新聞紙面制作のためのIT

インフラとは目的が異なる。「東スポ」の見出しや記事はひと目で「東スポ」と分かる独特の切り口を持つ。競合する新聞が事実上ない分、同社のコンテンツは人気が高い。東スポは、国会図書館でも閲覧頻度はトップクラス。「ぼろぼろになっているらしい」と吉良部長は笑う。そんな紙面には、「政治や経済や、いわゆる大事件の報道では伝わらない庶民の興味のあることを示す資料として価値があるのではないか」と吉良部長は言う。この計画にはこんな思いもこめられていた。

秒速で反応「新世代」のXMLデータベース

同社は、新しい紙面データベースの導入にあたり、印刷を委託している印刷会社の新聞紙面制作システムを手掛けた東芝ソリューションに依頼することに決めた。東芝ソリューションは、新聞紙面制作システムに実績があり、新聞社のニーズを理解していると考えたからだ。「ただ紙面をストックするだけの器を作っても意味がない。付加価値のある提案がほしい」と吉良部長は期待をこめて東芝ソリューションにこう伝えた。それに対して東芝ソリューションは、2つの提案を用意した。ひとつは、XMLデータベース「TX1」の導入だ。XMLデータベースは近年改良が進んでいる。東芝ソリューションの「TX1」もそうした新世代のXMLデータベースだ。約100ギガバイト（8年分の特許公開情報、約300万件に相当）のデータから、1万件の検索結果を抽出するのにかかる時間は1秒。すでに稼動していた、記事と画像それぞれのデータベースがXMLデータベースだったので、その有効性は実感していた。そして、もうひとつの提案が、東芝ソリューションが有料音楽配信サービスのシステム構築で培ったノウハウをもとに、紙面をインターネットで有料配信する仕組みだった。吉良部長は「すぐに使えるわけではないが、商用サービスの可能性を示してもらえた」と新聞業界を知る東芝ソリューションの提案を評価する。



XMLデータベースサーバ(MAGNIA 3415R)

※1 サムネイル／広告・出版業界の制作過程でイラストやコピーなどの配置を簡単に視覚化したもの。ここでは、多数の画像データを画面上に一覧表示するため縮小した画像のことを指す
※2 デジャブ／DjVu。従来の画像圧縮技術よりも小さなファイル容量に圧縮する新しい技術。米国AT&Tにより開発され、画像データ内の文字の可読性を損なわず、PDFやJPEGの5～1000倍の圧縮率を実現する

情報システム部
部長
吉良 輝夫 様



東スポとIT活用

新聞記者にとって「使える」感覚検索とサービス

東芝ソリューションの提案を受け、2005年9月、同社は「TX1」をベースに紙面データベースの設計・構築に乗り出した。この計画にはインターネット時代を見据えたデジタルデータの整備という目的と同時に、別の目的があった。独自の紙面となる同社の場合、他に参考になるものが少なく、自社の過去の記事にあたる機会が多いのだ。「記者が感覚的に使えるデータベースにしなければ意味がない」と吉良部長は当初から考えていた。しかし、言うは易く行なうは難しだ。記者によって、記事の記憶のしかたはまちまちだ。紙面上の位置で覚えている、隣の記事の大きな見出しだけを覚えているなど、あいまいな場合も多い。単純なキーワード検索だけでは、ほしい記事を探し出すのは難しい。そこで、紙面をそのままサムネイル^{*1}で表示し、文字情報と視覚イメージのどちらでも検索できるようにした。この解決策は、東芝ソリューションが同社の要望に応えるためにこらした工夫のほんの一部だ。「技術的なことには、それほどこだわりはない」と同部長は明かす。ただし、このデータベースに何が必要か考えぬき、必要と判断したことを実現するには妥協しなかった。さらに、毎日、大量の紙面データが登録されるデータベースへの負担も考慮し、画像データをコンパクトに圧縮する「デジャブ^{**2}」というデータ圧縮技術を採用した。「紙面全体のデジタル化に向け、圧縮率の高さと画質や文字の可読性を両立させてほしい」（吉良部長）という同社のニーズは、かなえられた。

現場主義の東スポとXMLデータベースの使い方

新しい紙面データベースは2006年3月の本番稼働を前に、1月4日の新聞からストックがはじまっている。ひとまず今後5年間分の150万件の蓄積を予定している。当面は、主に紙面づくりのために新聞記者が使う。便利な機能に、記者たちからは早くももっと前の紙面もデータベースにいられて欲しいという要望が出ている。ただし、

吉良部長は、記者たちにとって便利になり過ぎることへの警鐘を鳴らすことを忘れない。「新聞記者の仕事は、現場に行くこと、という基本精神は忘れないで欲しい」（吉良部長）。一方、将来的なインターネットでの有料コンテンツ配信計画に関しては、新聞各社とも有料コンテンツビジネスを確立できていない今、慌てて取りかかるつもりはない。しかし、インターネット向けに記事を再編集することは考えていないと断言する。「我々は毎日の紙面づくりに全ての力を注ぎ込んでいます。レイアウトもこれしかないというものだから他の形で外に出すことはない」という吉良部長の言葉には、新聞という形になった「東スポ」に対する自負がにじむ。ウェブコンテンツでも、独特の見出しや大きな写真で読者の心を惹きつけてやまない東スポらしさを、いかに出すか。そのユニークなコンテンツ配信が実現するときには、新聞紙面制作システムで、そして今回の紙面データベース構築で「東スポ」の紙面づくりを支える東芝ソリューションの役割はさらに大きなものになっているかもしれない。

Solution Focus

新聞トータルソリューション コンテンツデータベースシステム

素 材管理ソリューション、新聞制作ソリューション、クロスメディアソリューションを柱とし、多様なニーズに自在に対応できる“軸”の通った新聞トータルソリューション。このソリューションの一部であるコンテンツデータベースシステムが、記者が記事やイメージを高速に検索できるXMLデータベースシステムだ。紙面をそのままサムネイルで表示し、記者が文字情報と視覚イメージのどちらでも検索できるようにした。

<データベースエンジンに「TX1」を採用>

「TX1」は、XMLデータから構造を自動的に抽出して索引化する、構造自動抽出技術「スキーマアナライザ」と、抽出した構造情報と語彙情報を統計的に分析して最適な問い合わせプランを生成する、問い合わせ最適化技術「クエリオプティマイザ」を搭載することにより、大容量データでも高速に検索することができる。

この記事内容は2006年1月にNIKKEI NETにて取材した広告記事内容を元に構成しています。記事内における数値データ、組織名、役職などは取材時のものです。

●お問い合わせ先

東芝ソリューション株式会社

経営企画部 広報担当

住所：〒105-6691 東京都芝浦1-1-1(東芝ビルディング)

電話：03-3457-4112 MAIL：PR@toshiba-sol.co.jp URL：http://www.toshiba-sol.co.jp/jirei/

東芝ソリューション株式会社

