

トピック

・蒲郡市民病院におけるリアルタイム DWH 構築手法と BI における SDM 活用への展望

・「統計検定」のすすめ

ニュースレター第二号を発行させていただくにあたり、読者の方々から、できるだけ SDM の利用者の事例を紹介してほしいとのご意見をいただきました。本号におきましては、SDM 導入にいたるまでの経緯に関しまして、蒲郡市民病院の導入事例をご紹介いただけることとなりました。また、SDM に保存されたデータは活用することにより、導入のメリットが得られます。そこで、データ・サイエンティスト育成のための基礎知識である統計学に関する情報を掲載させていただきました。

一般社団法人 SDM コンソーシアム事務局



蒲郡市民病院におけるリアルタイム DWH 構築手法と BI における SDM 活用への展望(前編)

飯田 征昌

今回の寄稿に際しましては、2021年11月に開催されました第41回医療情報学連合大会のランチョンセミナーにおける発表内容に基づき、当院が2021年5月に実施した病院情報システム(以下 HIS)の全面更新において取り組んだ DWH 構築と BI の活用に関しまして、診療情報を基軸とした院内データの集約と活用の理想像、および SDM 活用に向けてのアプローチについての考察を、次号との2回に分けてご紹介いたします。

1. なぜ蒲郡市民病院は HIS を全面更新したのか?

先ず当院の概況についてご紹介します。当院は、2018年に名古屋市立大学と寄附講座設置に関する連携協定を締結し、同年4月に着任した城 CEO のリーダーシップの下、医師の増員やこれに伴う診療・研修・教育などの機能強化などを図ってきました。その結果、翌年度には開設以来初の黒字化を達成するなど、名実ともに大学との機能連携のメリットを活かした施策を実践してきました。

そのような流れから、やがて更新時期を迎える本院の HIS のあり方についても検討された結果、名古屋市立大学病院で築き上げてきた診療情報の一元管理に基づいたデータ利活用モデルを導入し、大学と一体となったデータ利活用の促進を通じて、DX 推進や地域包括医療システム構築など、今後の重要な成長戦略を

担う基盤としての HIS 更新計画の推進に至りました。

結果、従来の電子カルテ(以下 EMR)パッケージシステムであった富士通・EGMAIN-GX から、名古屋市立大学病院が独自に構築を行ってきた EMR である NeoChart への移行を中心とした HIS 更新を2019年11月に決定し、2021年5月に新システムの稼働を迎えることとなりました。

2. NeoChart アーキテクチャの特徴と変遷

今回導入した NeoChart のデータ集約に関するアーキテクチャの特徴についてご紹介する前に、僭越ながら私の経歴についてご紹介いたします。

前職は、先に挙げた名古屋市立大学病院において、1999年より HIS の開発・運用に関する業務に2019年の異動まで一貫して従事し、その翌年の退職を経て2020年11月より現職に就いております。

その間において、当時の HIS ベンダーと共同で NeoChart の様々な要件定義・プロトタイプモデリングに始まり、その後の運用におけるプログラム修正・開発やインフラ設計・監視といったエンジニアリングからマネージメントまでの課題解決に関する全般、システム更新の企画・監督など

「HIS に関する業務のあらゆる経験したのでは?」と述懐するに至ります。

そのような経験の下で最も

重視したことは、データの利活用を促進するためのデータモデリングであり、如何にして日々積み上げているデータとアプリケーションの基盤を陳腐化させずに有効活用し続けられるかという点に腐心しながら、現在に至る NeoChart のデータ基盤に投影してきたつもりです。

続いて、NeoChart 構築当初における基本的なデータベース設計モデルにおける特徴を通じて、その後の利点活用と弱点克服のアプローチについてご紹介したいと思います。

2.1 ドキュメント型データの利点を活かした診療情報集約

NeoChart は、記録やオーダーなど診療情報の主要部を格納するデータベースの設計において、一般的なリレーショナルモデルを採用せず、ドキュメント型のモデルを採用(最小限のキー・インデックスとドキュメント(XML)形式のデータで構成)を採用した結果、以下の特徴を有しています。

- ・利点: データの内容・構造に左右されず、多種多様なデータ格納が可能なため、アプリケーションのカスタマイズや新規開発が容易

- ・弱点: 最小限のインデックスで絞り込めないデータは、格納されたドキュメントを全て展開して探索する必要が生じ、特定要素(薬品・検査項目など)を指定した患者横断的な検索を高速で行うことは不可能

目次

蒲郡市民病院におけるリアルタイム DWH 構築手法と BI における SDM 活用への展望(前編)

飯田 征昌 1

「統計検定」のすすめ

本多正幸 3

SDM Tips 2

鈴木英夫 4

事務局より 5

よって、診療情報 DWH が構築されていない状況において、EMR のデータ抽出を行うにあたっては、SQL クエリ実行+XML ドキュメントの解析・正規化が必要となり、その処理プラットフォームを MS Access を用いて独自開発することで、医療従事者からのデータ抽出依頼や簡易アプリケーション開発などに用いてきました。

そのような状況下から、病理や読影レポートなどの見落としや、B 型肝炎再活性化のリスクマネジメントなど、医療安全面の課題解決を図る EMR 機能開発を求められた経緯を通じて取り組んだのが、部門システムで扱われる診療情報の集約・一元化です。

この取り組みにおいて、先述のデータベースモデルの利点が大いに活かされました。病理、内視鏡、読影、生理検査、眼科、分娩、手術、重症、救急の各部門システムで発生したカルテ記載に相当するデータ全てを、各ベンダーに XML で出力するよう依頼し、受領したデータそのままの状態を保持しながら NeoChart で入力した診療記録と完全に同等の扱いでの取込を実現しました。まさに、ドキュメント型モデルの利点をそのまま活かした形としています。

また、その連携基盤の構築においては、高度な処理プロセスをリアルタイムで広汎に行う必要性とその重要性を鑑み、SOA プラットフォーム

(InterSystems Ensemble) を採用し、プロセスの可視化・集約による安全性と効率性の両立も図りました。

上記を通じて、部門システムの診療記録は EMR そのものの記録としてリアルタイムに格納されることから、各種レポートの内容までもが同列に参照できることを始点として、未読通知や既読管理のマネジメントから、診断結果に応じた次の臨床プロセスの提示といった臨床意

思決定支援システム(以下 CDSS)に相当する機能なども、EMR 上で一元的に実装可能なデータ基盤として構築しました。

2.2 ドキュメント型データの弱点击破のためのリアルタイムマルチデータモデル実現

しかしながら、集約したデータの活用の際には従来からの弱点が課題となります。先述の B 型肝炎再活性化における CDSS を例に、

「免疫抑制剤など特定薬剤の投与指示をリアルタイムで補足し、肝炎に関する検査結果の直近の値を判定した上で、更に肝炎に関する検査の今後の指示状況をモニタリングして判断した上で、検査指示が必要であれば介入する」といった場合を挙げます。

従来のドキュメント型モデルのみのデータ検索では、数十秒～数分以上の検索時間となった実績からも、瞬時のサジェストは事実上不可能です。その一方で、この様な重要性の高い介入行為こそ医師が指示したそばから行えなければ、有用ではないとも考えられます。

よって、この課題を解決するために取り組んだのが、ドキュメント型データとして格納されている様々なデータ要素をインデックスとして利用可能にし、その生成はリアルタイムに行う施策です。先に、その手法の概要を図に示します。

本手法の特徴は、その処理全てをデータベースサーバーのプログラム処理で完遂させる点です。その目的と最大の利点は、従来の EMR アプリケーションに一切の修正を加えずとも、検索に適したデータを生成できる点であり、当初の設計の利点を活かし続けながらコストも最小限に抑えられる手法と認識します。その実現に際しては、従来の RDBMS からマルチモデル DBMS (InterSystems Caché) の採用を前提とし、サーバー側の

プログラマブルな処理機能をフル活用するアプローチとしました。

結果、ドキュメント (XML) として格納されたデータ集合の発生に連動して、データ集合を構成する最小単位のデータを格納するシンプルな Key-Value Store (KVS) をリアルタイムに生成するモデルとし、ドキュメント単位とデータ要素単位の 2 つの KVS を同時に構築することにより、

- あらゆる構造のデータを柔軟に格納可能
 - あらゆる視点からのデータ抽出の高速化が可能
- を実現することによって、先述の CDSS に必要なクエリを迅速 (数秒以内) に実現可能な EMR データベースの構築を、2015 年頃より実践し現在に至ります。

当初、サーバー側においてトランザクション発生時のデータロード・解析といった処理プロセスを要求することから負荷面での懸念はありましたが、ハードウェア更新サイクル単位で見た場合のワークロード変化に比して、CPU・メモリ・ストレージの性能向上の度合は圧倒的であり、その範疇で十分に収まっているものと評価します。

データ必要時点でのフルパワー要求や、別のコンピュータリソースありきでは無く、その余力を活かして格納するそばから必要時に備えておく合理性こそが、迅速・広汎なデータ利活用の礎であると考えます。

今回は、リアルタイム DWH 構築の前提としての、EMR の機能性向上のためのデータベースアーキテクチャの変遷についての紹介のみにて紙面が尽きてしまいましたが、次回は先述の KVS を DWH のデータソースとして利用しながら BI ツールの応用に繋げる手法の実例を通じて、SDM 活用への展望について考察します。



著者 飯田 征昌

1980 年 愛知県生まれ
 2007 年 名古屋工業大学工学部卒業
 1999 年 名古屋市立大学病院入職 病院情報システム担当、保険・診療情報管理担当主査を経て
 2020 年 10 月 退職。
 2020 年 11 月 より蒲郡市民病院 事務局
 デジタル化推進監
 上級医療情報技師

Caché のプログラマブルな機能を活用し、XML データ更新時のトランザクション内で項目要素とデータを抽出し、データ項目毎の Key-Value Store 方式でリアルタイムに格納する処理を追加

Patient ID	Order Number	Date	State	BinaryStream	Array_Key	Array_Value
3000	123456	2013/8/28	0	<?xml version="1.0"?> <PatientID>3000</PatientID> <OrderNumber>123456</OrderNumber> <Prescription> <RP Num="1"> <DrugGroup> <DrugCode>1111</DrugCode> <DrugName>ガスター	DrugCode	1111
					DoseCode	246246
					Update UserID	iplida
				

ほとんどの情報が XML 文字列の集合であり、特定の要素を指定して高速に検索することは困難

処方、検査結果、テンプレート記録などの検索において、
 ・特定項目に絞った時系列の検索
 ・特定項目の値の条件を指定した抽出
 など、従来不得手の検索条件における飛躍的な改善を実現

元データを最小粒度に分解して格納

「統計検定」のすすめ

1. はじめに
第四次産業革命 (Fourth Industrial Revolution, 4IR)は18世紀の最初の産業革命以降の4番目の主要な産業時代を指すが、この革命はロボット工学、人工知能 (AI)、ブロックチェーン、ナノテクノロジー、量子コンピュータ、生物学、モノのインターネット (IoT)、3Dプリンター、自動運転車などの多岐に渡る分野においての新興の技術革新が特徴である。その中で著者は、AIやIoT、ロボティクスに関心があるがそれらの先進的な展開は、すべてビッグデータと呼ばれる膨大なデータ群とそのデータを処理できる情報技術の革新に大きく依存していると考えている。また、AIと呼ばれている分野では、人間の思考、判断、予測などを模倣するための技術の総称であり、それらを実現する一つの手法として機械学習があり、更にニューラルネットワークを用いて学習を効率よく行う機械学習がDeepLearning(DL)である。医学や医療の分野では、ゲノム医療、画像診断支援、診断・治療支援、医薬品開発などの分野において積極的な投資が行われつつある。画像診断支援では、日本医療研究開発機構 (AMED)において「医療ビッグデータ利活用を促進するクラウド基盤・AI画像解析に関する研究」を数年前より立ち上げて成果を出している。このプロジェクトは、日本医学放射線学会などの医用画像を取り扱う医学系の6つの学会と国立情報学研究所 (NII)において、膨大な医療画像を蓄積し、蓄積されたデータに対しAI研究者らがそれぞれ提示した課題を解決しようとしたオールジャパンのプロジェクトである (残念ながら令和3年度で最終年度となっている)。
本稿では、機械学習における統計学との関連を実例と共に紹介し、AIといっても基盤となる技術は、我々が学習してきた情報科学の技術に他ならないということである。情報科学の中でここでは統計学に焦点を当て、

基盤技術である統計学を改めて学習することの意義を確認したい。最後に統計学を学習する際の確認手法の一つである「統計検定」の存在を紹介するとともに、「統計検定」の概要を示す。過去に行われた試験問題は、統計検定のHPに掲載されているので、各自で各級のレベルを実感していただき、積極的にチャレンジする方が一人でも増えることを期待したい。
<https://www.toukei-kentei.jp/past/>

2. 機械学習と統計手法
機械学習には、大きく分けて以下のように3通りに分類される。

- 教師あり学習・・・正解がわかっている場合の学習で、AI (人工知能)が問題を解き、正解との誤差が少なくなるようにAIの挙動 (内部パラメータ)を修正していく方法。
- 教師なし学習・・・正解がない場合の学習で、問題を観察し、似かよった問題を分類し、まとまり (グループ)を発見する方法。未知の問題に対してどのグループに属するか認識し、対応する方法。
- 強化学習・・・試行錯誤を行いつつ最適なと思われる「行動の仕方」 (方策、ポリシー)を学習する方法。(教師あり学習で正解が少ない場合には、「半教師あり学習」「転移学習」「ゼロショット学習」などがあるが、本稿の範囲ではないので割愛する。)

機械学習で利用されている統計手法としては「教師あり学習」では回帰分析 (最小二乗法)があり、分類問題などにも適用されている。スパム問題などに適用される単純ベイズ分類器では、確率論で勉強するベイズの定理が利用される。また、「教師なし学習」では、主成分分析やクラスタリング分析などの手法が適用されている。このように各種の統計手法がAI分野で活用されており、再度統計学や統計手法を見直してみるこ

本多正幸

とは時流に合った課題ではないだろうか。

統計検定 (JSSC) ホームページ <https://www.toukei-kentei.jp/> における紹介文は以下のようになっている。

- 「統計検定」とは、統計に関する知識や活用能力を評価する全国統一試験です。
- データに基づいて客観的に判断し、科学的に問題を解決する能力は、仕事や研究をするための21世紀型スキルとして国際社会で広く認められています。
- 日本統計学会は、中高生・大学生・職業人を対象に、各レベルに応じて体系的に国際通用性のある統計活用能力評価システムを研究開発し、統計検定として実施します

試験の方式は、PBT (紙テスト) と CBT があり、試験の種類は、1級、準1級、2級、3級、4級と分かれています。またこれらとは別に、統計調査士、専門統計調査士があり、最近加わったものにデータサイエンス基礎およびデータサイエンス発展があるが、1級以外はすべて CBT として2022年以降実施される予定と聞いている。CBTとは、コンピュータを用いて受験する方式で、全国で日時や会場を選んで各自に合わせた学習計画を立てやすいといったメリットがある。

表1は、各級における試験内容とその目安を示している。個人的な見解・希望であるが、医療情報関係者並びに医療情報技師、診療情報管理士の方々には、ご自身の統計・統計学の知識を把握するうえで本統計検定の活用をお勧めしたい。大学病院での情報システムの管理に関する経験や、医学部及び看護学校での統計学の講義経験から、読者の皆様においては2級を目指していただきたいが、まずは3級合格が当初の目標としていただければと考える。このニューズレターの読者は幅広い層の方々かと推察するが、この仕組みを活用されてキャリアアップへと繋げていただければ幸いです。

表1

試験の種類	試験内容	レベルの目安
1級	実社会の様々な分野でのデータ解析を遂行する統計専門力	大学での専門分野修了程度
準1級	統計学の活用能力 ー 実社会の課題に対する適切な手法の活用能力	大学卒業程度
2級	大学基礎統計学の知識と問題解決力	大学基礎課程程度
3級	データの分析において重要な概念を身に付け、身近な問題に活かす力	高等学校の「データの分析」程度
4級	データや表・グラフ、確率に関する基本的な知識と具体的な文脈の中での活用能力	中学卒業程度
統計調査士	統計に関する基本的知識と利活用	経済統計 / 社会統計 / 公的統計 対応
専門統計調査士	調査全般に関わる高度な専門的知識と利活用手法	2級の知識+広範な能力
データサイエンス基礎	具体的なデータセットをコンピュータ上に提示して、目的に応じて、解析手法を選択し、表計算ソフト Excel によるデータの前処理から解析の実践、出力から必要な情報を適切に読み取る一連の能力	AI・デジタル社会の共通スキルを評価
データサイエンス発展	数値・データサイエンス教育強化拠点コンソーシアムのリテラシーレベルのモデルカリキュラムに準拠した内容	
データサイエンスエキスパート	数値・データサイエンス教育強化拠点コンソーシアムの応用基礎レベルのモデルカリキュラムを含む内容	

キーワード：
AI (人工知能)
機会学習
統計学
統計検定

著者 本多 正幸

長崎大学名誉教授
(平成31年3月31日退職)
千葉大学特任教授
(令和2年6月 企画情報部)

職歴：千葉大学附属病院および長崎大学病院にて病院情報システムの管理の仕事を行ってきた。大学では医学統計に関するコンサルテーションや講義も担当し、現在「統計検定」に関する役員も兼任している。

SDM Tips No.2

鈴木英夫

前号では、なぜ現在の DBMS の主流が RDB であるのかを理解するために、RDB の歴史について紹介しました。今回は、Semantic Data Model に関して紹介します。

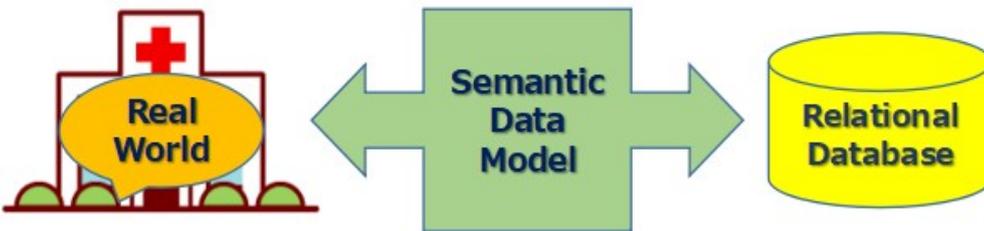
データベースモデルは、物理ストレージ上にデータを保存するための手法のことを言い、階層型データモデル、ネットワーク型データモデル、リレーショナルモデル、オブジェクト・データモデルなどがあります。これらは、実世界で発生するデータ（リアルワールドデータ）を物理ストレージに保存する際に用いられる手法ですが、どの手法を用いたとしても、リアルワールドデータの概念的定義を完全に表現することはできません。そこで考案されたのが、Semantic Data Model です。このモデルは、リアルワールドと RDB 間の関係を定義する概念モデルであることから、概

念的ビューとも言います。Semantic Data Model では、データベース上で、データ間の関係性、およびデータの意味を完全にリアルワールドと一致させることが必要となります（図）。

実は、後で知ったことですが、この Semantic Data Model が最初に用いられたのは、1976 年 US Air Force Materials Laboratory の ICAM(Integrated Computer-Aided Manufacturing)プログラムと言われています。筆者も IBM の基礎研究所では、肺癌の CAD(Computer-Aided Diagnosis)を主研究テーマとしており、開発したアルゴリズムも、赤外線探索追尾装置で目標物を検出するためのアルゴリズムからヒントを得たものでした。

これらの目的志向型研究においては、その目的に必要なデー

タを取得する「前向き分析」となります。一方、現在ではセンサー技術やデータ保存技術の進歩により、取得可能なデータの種類やデータ保存量が格段に増えており、目的を特定しない、あるいは別の目的で取得・保存されたデータを利用して新たな知見を得る「後ろ向き分析」でも、十分な成果が得られるようになってきました。電子カルテにおける臨床データは、患者の診療記録として保存されていましたが、現在では、臨床研究、臨床支援、臨床試験、臨床教育、疫学調査、業務改善、経営分析、など様々な目的で二次利用されています。しかし、二次利用のためには一次目的で保存されたデータを様々な視点で検索できるデータベースモデルが必須となります。リアルワールドにおいて発生するデータそのものは、「符号」であり「情報」としては不完全です。



例えば、電子カルテで病名を登録する場合、電子カルテの画面上では、病名という表示とデータ入力域、および実際の病名の選択リストなどが「情報」として表示されていますが、保存されているのは、コードという「符号」だけであり、「情報」として復元するためには、その「符号」が病名コードであるという定義と、そのコードが何の病名であるという変換表(マスタ)が必要になります。この変換表に関しては、RDB における関係モデルにより、復元が可能となりますが、個々のデータの意味に関しては、項目名だけでは、元の「情報」への復元は困難となります。例における、「病名コード」という項目名も、どのパー

ジョンのどのコード体系かが不明だと、病名への復元は出来ません。したがって、リアルワールドとデータとを関係付けるための Semantic Data Model が必要となるのです。

この Semantic Data Model の考え方を拡張したモデルが SDM です。SDM では、普遍的な項目を共通項目としても、特定の分野で利用されている項目を、その分野における用語として採用しています。普遍的な項目とは、行為を説明する項目群「何時から何時まで」「何処で」「誰が」「どの立場で」「誰のために」「何を」「どうした」などです。

例えば、「2022 年 1 月 1 日 午前 8:00 から 2022 年 1 月 1 日 午前 8:10 まで、病棟

A において、内科の患者 B のために、チームの看護師 C が、患者 B の検温を行い」ここまでが共通項目となり、「体温が 36.5 度であった」はテーブル固有項目として保存されます。

この普遍的な共通項目を設定したことにより、様々な切り口での検索を可能とすることが出来ました。

次回から、SDM の個々の特徴とメリットについて説明していきます。



【 SDM 】

Semantic Data Model

【 RDB 】

Relational Database

【 DBMS 】

DataBase Management System

【 CAD 】

Computer Aided Diagnosis

【前向き分析】

【後ろ向き分析】

著者 鈴木英夫

株式会社 M o D e L 代表取締役

出身

神奈川県横浜市

学歴

千葉大学工学部

千葉大学大学院工学研究科

千葉大学大学院自然科学研究科

(知識を利用した胸部 X 線画像処理とその応用: 学術博士)

職歴

IBM Japan (1983-2014)

IBM Corp. (1998-1999)

(CIS-Image, MDView を開発)

MoDeL Inc. (2014-)

(SDM を開発)

SDM Consortium (2014-)

Findex Inc. (2017-2020)

教歴

千葉大学非常勤講師

東京女子医科大学非常勤講師

岐阜大学非常勤講師

鳥取大学非常勤講師

千葉大学客員准教授

一般社団法人
SDM コンソーシアム
郵便番号 223-0066
横浜市港北区高田西 2-4-10

電話
(045) 567-3613

FAX
(045) 567-3613

電子メール
info@sdm-c.org

ユーザーが育てる DWH

活動記録(2021年)

【SDM データベース定義書】

V1.13 6月9日

HPよりダウンロード可

【第8期】

2021年4月1日より
2022年3月31日まで

【社員総会】

5月26日

【理事会】 4回

【役員会】 月2回

【懇談会】 月1回

【学会参加】

MTA 学会大会 8月

SDM(Semantic Data Model)について

SDMは、ヘルスケア情報に関するDWH(データウェアハウス)の設計書です。DWHの設計においては、項目間の関係を含めた構造(モデル)が必要となります。項目の意味(Semantics)を理解した上でその関係を構築する手法

(Semantic Data Modeling)を用いることにより、有意義な2次利用が出来るようなDWHを構築することができます。

これがSDMと命名した理由です。

SDMコンソーシアムは、SDM設計書の整備・保守・分析サンプル作成・公開・アナリストの教育・スキル育成、プロジェクトの支援、セミナー・ミーティング・イベントの企画などを通して、SDM普及を行っております。

一般社団法人 SDM コンソーシアム
〒 223-0066
横浜市港北区高田西 2-4-10



組織・役員 (2022年1月1日現在)

【代表理事】	紀ノ定保臣	(岐阜大学)
【監事】	本多正幸	(長崎大学・千葉大学)
【理事】	近藤博史	(鳥取大学)
	久島昌弘	(沖縄県立中部病院)
	村垣善浩	(東京女子医科大学)
	木村映善	(愛媛大学)
	島川龍載	(県立広島大学)
	飯田征昌	(蒲郡市民病院)
	鈴木英夫	(株式会社 MODEL)

【事務局】 濱田麻里

【賛助会員】 富士フィルムメディカル IT ソリューションズ株式会社
株式会社医用工学研究所
株式会社ジャストシステム
株式会社ファインデックス
株式会社医用ラボ
株式会社ソフトウェア・サービス
NCS&A 株式会社
IQVIA ソリューションズジャパン株式会社
コスモ開発株式会社
メディカル・データ・ビジョン株式会社
株式会社エムケイエス
データキューブ株式会社
株式会社アドバンスト・メディア

Web サイトにてお待ちしております。

Web サイト アドレス:
<https://sdm-c.org>