

CoreExplorer Ver.3.0 のリリースと今後の事業展開

Future Business Deployment of the Text-Mining System CoreExplorer

(株)日立ソリューションズ東日本(HSE)では、テキストマイニング製品である CoreExplorer のソリューションを展開し約 11 年となった。これまで、製造業のお客様を中心に、公共・文教、流通・小売、サービス、金融・保険など様々な業種で製品導入、コンサルティングを行い、経験を積んできた。近年ではソーシャルメディアやビッグデータから知見を発掘し、事業拡大や業務改善に繋がりたいと考えている企業がテキストマイニングへ関心を寄せており、それに伴って市場も拡大している。

本動向を踏まえ、CoreExplorer は 2014 年 5 月にメジャーバージョンアップを行い、CoreExplorer Ver.3.0 をリリースした。今後は更なる製品力強化と日立グループ内の連携により、事業拡大を図っていく予定である。

宮林 靖典	Miyahayashi Yasunori
宮内 秀彰	Miyauchi Hideaki
塚原 朋哉	Tsukahara Tomoya
川原 悠	Kawahara Yuu
佐々木 智和	Sasaki Tomokazu
佐藤 俊也	Sato Shunya

1. はじめに

近年、企業内に蓄積されている大量のテキストデータを活かして新たなニーズや対処すべき問題点を発見し、迅速な意思決定とアクションに繋げていくことは、ビジネスを成功させる上での重要な要素となっている。

製品の製造過程で発生した故障品の情報や、市場で発生した不具合の情報から、問題の早期解決や事故の防止に役立てるといった利用方法や、コールセンタに寄せられるお客様の声などからニーズの抽出やクレームの要因分析を行うなど、重要な情報がテキストデータに存在していることは多い。

これらの故障・不具合情報や、お客様の声など、自由記述文で蓄積されているデータは、数値情報のように簡単に集計・傾向把握することが難しい。

こういった課題に対し、HSEではテキストマイニングシステムであるCoreExplorerを開発し、短時間で定量的にテキスト分析を行える製品として市場に提案してきた。

本稿では、事業開始から現在までのCoreExplorerの事業実績、およびお客様の評価や市場の動向を基にリリースしたCoreExplorer Ver.3.0の紹介、さらに今後のエンハンス計画と、事業展開の方針について述べる。

2. テキストマイニング事業の取り組み

2.1 CoreExplorer事業の歩み

2003年にリリースしたVer.1.0から現在のVer.3.0に至るまで約11年間事業を継続中である。

その間CoreExplorerを活用し、業種・業態にとらわれずソリューションを展開してきた。例えば、製造業の品質保証部門における品質分析や、カスタマーサービス部門でのお客様の声分析、社内に蓄積された文書の検索、人と技術の関連性を可視化したKnowWho検索など、様々なソリューションに展開している。

大学や日立グループ内の研究所とも積極的に研究交流を図っているほか、近年では日立グループ内外の製品との連携実績も増えてきている。

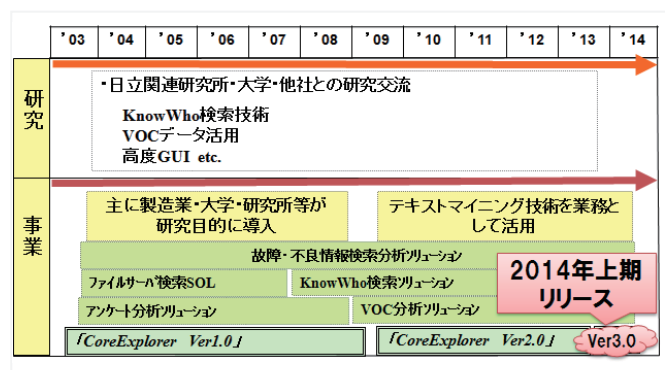


図1 CoreExplorer事業の歩み

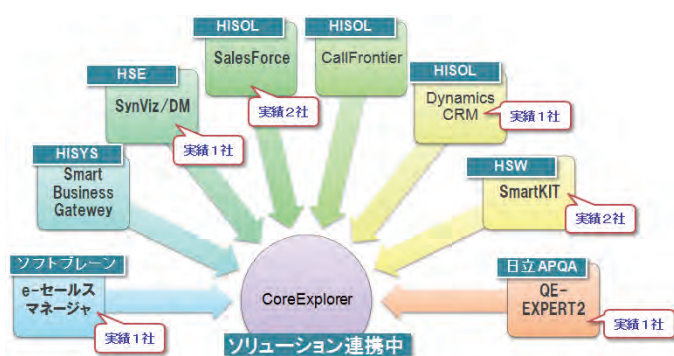


図2 CoreExplorerの主な連携製品

2.2 CoreExplorerの導入実績(2003年～2014年)

CoreExplorerのこれまでの導入実績を表1に示す。

CoreExplorerがターゲットとする製造業のお客様は64社中36社と全体の半数以上を占めているが、その他の公共、流通、金融といった分野でも導入実績があり、幅広い業種・業態で採用されている。

表1 CoreExplorer の導入実績

業界	主な用途	件数
製造業	故障報告・不具合報告の検索、分析、自動分類	15
品証部門	故障報告・不具合報告の検索	4
設計部門	故障報告・不具合報告の検索	4
企画部門	お客様満足度調査分析、Q&Aシーートの分析、VOCの自動分類や分析	2
営業部門	営業報告の分析	2
その他	特許検索、コールセンタ情報分析、営業情報分析	13
公共・文教	KnowWho検索、FAQ検索、アンケート分析、口コミ分析、医療情報(ヒヤリハット)の分析	15
流通・小売	お客様からの要望、クレーム分析	2
サービス	アンケート分析、調査情報分析	5
金融・保険	コールセンタでの問合せ分析、メールコンテンツの検索、SFAの分析	6
合計		64

3. CoreExplorer Ver. 3.0 のリリース

3.1 CoreExplorer Ver. 3.0の開発動機

CoreExplorerを導入されたお客様からはこれまでに以下のような要望が寄せられていた。

- ①一度の分析で取り扱えるデータ件数を増やして欲しい
- ②日本語以外の言語でもテキストマイニングしたい
- ③見落としていた有益な情報を得やすくするための分析機能を増やして欲しい

これらはお客様が業務を進める上で特に重要度の高い要望であった。

3.1.1 大規模データへの対応（数十万件規模）

サーバOSの64bit化に伴い、近年企業内の1システムが扱うデータであっても数十万件規模に達することも珍しくない。テキストマイニングの対象となるデータも、数千件、数万件のオーダーから数十万件のオーダーに変わってきている。

例えば、大規模コールセンタのお客様では1日に数千件から数万件程度の問合せがあり、分析対象の問合せ履歴データは月あたり数十万件に達する。

Ver.2.0は、32bitOS上で動作するため、CoreExplorerの検索時にキャッシュできるメモリ容量に2GBの制限があった。そのため検索性能を考慮した場合、数万件のデータしか扱うことができなかった。

Ver.3.0では、大規模データを扱えるようにするために、まず64bitOSに対応することで大規模データを扱えるプラットフォームにしておく必要があった。

3.1.2 多言語の分析に対応

近年日本の企業は海外へビジネスの範囲を拡げている。例えば製造業の企業では、海外に工場を持つことが多く、蓄積される故障情報などは現地の言語で書かれている。また、製造業各企業の販売先も海外の割合が増えており、日本語圏以外のお客様の声も集まってきている。こういった背景により、日本語以外の言語の解析技術が求められている。

今までは日本語解析を専用とするシステムであったが、海外展開している国内製造業の企業をターゲットとするために分析言語を多言語化する必要があった。

3.1.3 ビジュアル機能強化による分析機能の拡充

HSEではこれまで、様々なお客様に対してCoreExplorerを利用したコンサルテーションを行って

きた。その中で、分析対象の文書から見落としていた有益な情報を得るための分析作業をアシストする機能や、分析した結果をより適切に表現できるグラフ機能の要望を多く頂いた。

CoreExplorerが持つ円グラフ・棒グラフ・折れ線グラフなどの基本的なグラフ機能に加えて、お客様による分析をビジュアルでサポートする機能や、分析結果をより効果的に表現するグラフ機能を拡充する必要があった。

3.2 CoreExplorer Ver.3.0の新機能

前章で述べた要望に応えるべく、Ver.3.0では以下の新機能を実装した。

- ①64bitOS対応・性能向上
- ②英語、中国語のマイニング機能実装
- ③マップ機能強化・グラフ機能拡充

今までは分析データ件数の不足や、分析言語の不足により、大規模データを扱う企業や、海外展開を図る国内の大手企業に対して積極的なアプローチができない状態であった。Ver.3.0でこれらの課題が解決されることから、今後は積極的に提案し、受注に繋げていく。¹⁾

3.2.1 64bitOS対応・性能向上

前章3.1.1に挙げているように、お客様がテキスト分析の対象としたいデータは大規模化してきている。

CoreExplorerで扱えるデータ件数はこれまで1サーバでは32bitの制約によるメモリ使用量の制限や、データの更新・検索性能等の制限により、数万件規模であった。

Ver3.0では、検索エンジンの64bit化やメモリ使用効率化により、検索性能を向上させた。また、データ件数が増加した場合、検索インデックスを作成するデータ更新性能も日々運用する上で重要な課題であった。本課題に対しては、形態素解析処理の高速化や複数インデックス作成の並列処理等の対策により、データ更新性能を向上させた。これら対応により1サーバあたり数十万件での分析が可能とした。

また、内部的にはシングルサーバ構成ではなくクラスター構成に耐えうるアーキテクチャに変更し、今後データ拡張が生じて柔軟に対応できる構成を実現した。

3.2.2 英語、中国語のマイニング機能実装

前章3.1.2に挙げている、分析言語に対する要望に対し

て、Ver.3.0ではOSSである汎用形態素解析器『Mecab』を利用し、中国語・英語の解析機能を実装した。Mecabは、現在CoreExplorerの日本語形態素解析エンジンに利用している『Chasen』と比較して高速に動作し、かつ言語・辞書に依存しない汎用的な形態素解析エンジンである。そのため、辞書を追加することによって、様々な言語に対応することが可能である。Mecabを利用することにより、英語、中国語の分析を実装したが、今後の市場の動向に合わせて、更なる言語拡張にも対応できるアーキテクチャとした。

3.2.3 マップ機能強化・グラフ機能の拡充

前章3.1.3に挙げている、分析機能の拡充として、マップ機能のビジュアル強化・新たなグラフ機能の追加を行った。具体的には、マップに配置される単語間における関連の強さを補助線の長さで表現したり、より特徴的な単語をノードの色で強調したり、特徴的な単語を含む文書数をノードの大きさに表現するなど、ビジュアル機能を強化することで、分析をサポートする情報量を増やしている。また、データのばらつきや相関を分析するための散布図を実装した。

これらの機能により、性別や年代、地域など各属性やキーワードの因果関係・相関関係を得ることが容易となった。



図3 マップ表示の例

4. 今後の事業展開

4.1 今後の市場動向

テキストマイニング市場はビッグデータ活用の機運が高まっていることを受けて活性化している。ITRの調べでは、2012年度の市場規模調査では出荷金額が28億9,000万円で、前年比11.6%増と拡大しており、2013年度

以降も引き続き拡大基調にあると予想されている。

また、ITRでは、テキスト分析のニーズとしては、従来のアンケート分析やビジネステキスト分析に加え、ソーシャルメディア(Facebook, Twitterなど)の分析ニーズが高まっているとしている。²⁾

最近、HSEへもソーシャルメディアやビッグデータの活用を検討されるお客様からの引合いは多く、データソースはより大規模になっている。また、業務システム内に蓄積されている大規模のテキストデータを事業に活かしたいと考えている企業から、テキストデータの分析・活用方法を支援して欲しいとの要望があり、コンサルティングのニーズも高まっている。

昨今の安価に導入できるSaaSの提供や、コンサルティングサービスの強化など各社のサービスが多様化し、テキストマイニング市場はさらに競争が激化することが予想される。

4.2 Ver. 3.1でのエンハンス予定

Ver.3.0をリリースした後も引き続きお客様からは下記に示す要望が寄せられている。今後Ver.3.1にて要望に応える機能拡充を行い、製品の機能強化を行う予定である。

4.2.1 更なる大規模データへの対応

大規模コールセンタでは1日に数千件から数万件程度の問合せがあり、年間数百万件に達する。またTwitterなどのソーシャルメディアでは、特定のテーマで抽出したデータを分析する場合でも数百万件、雑多なデータを半年から1年程度の時系列で分析する場合は数千万件から数億件のデータを扱える性能が求められる。現在CoreExplorerで扱えるデータ件数は1サーバあたり数十万件までであり、競争力を維持していくために大規模データへの対応は重要な課題である。

4.2.2 データ取込みの簡易化への対応

CoreExplorerを導入しているお客様より、分析業務の効率を向上させるために、分析前のデータ取込み作業に掛かる負荷を軽減したいというニーズが高まっている。現在、CoreExplorerを含め分析系の製品はデータ取込みが手動である場合が多く、お客様から自動化への要望を頂いている。

これまでのデータベース連携に加え、データ管理系のシステムから抽出された CSV を自動で取り込むことが

できる機能を実装することで、CoreExplorerの利便性が高まると期待されている。

4.3 更なる事業拡大に向けて

HSEでは、今後もCoreExplorerによる更なる事業拡大をめざしている。事業拡大の要素としては、図4に示すように、大きく2つの事業内容を検討している。

4.3.1 API提供によるソリューション連携施策

まず、CoreExplorerのテキスト分析エンジンを部品化しAPIとして販売する事業である。

企業が利用している業務システムに蓄積されているテキスト情報を、テキストマイニングツールを意識せずにシームレスな形で利用したいとの要望がある。

その要望に対し、テキストマイニング技術をAPIとして活用することで、既存の業務システムや製品の使い勝手や付加価値を高めることができる。例えば、API提供により既存の業務システムや製品に次のような機能を追加することが期待できる。

システム内に蓄積されているテキスト情報を検索する際、入力した検索キーワードと関連が強いキーワードを表示することで、検索をアシストする機能である。ユーザの検索を支援することで、よりすばやく得たい情報に到達することが可能となる。

またアンケートや不具合情報を目視で確認し分類する業務では、テキストデータを分類する条件にしたがって自動的に分類を行う自動分類機能を利用し、既存システムのデータベースに分類結果を反映させることで、これまで目視で確認していたデータ分類業務の効率化が期待できる。

日立グループ内には、コールセンタシステムや、保安全管理システム、電子帳票システム、といったテキストデータが蓄積される管理システムが多数存在している。これら製品にテキスト分析の機能を付加することで、管理から分析まで一貫したソリューションの提供が可能となり、ソリューション競争力を高められると考えている。

4.3.2 価値あるアウトプットを提供するための施策

もう一つが、CoreExplorerを利用するお客様がより価値のあるアウトプットをより簡単に導くことができるように、使い勝手の向上と簡易に分析できる機能を拡充する施策である。CoreExplorerは販売実績が10年以上にも達し、基本アーキテクチャや利用技術も最新化する必要

があると認識している。今後は、現状のCoreExplorerで得意とする使い勝手の良さを、HTML5やCSS3といった画面デザインの最新技術によりさらに高めるとともに、仮説立案を支援しより簡単に分析できる機能を拡充していく。この施策により、テキストデータを業務に活用できる利用者の裾野を広げていきたいと考えている。

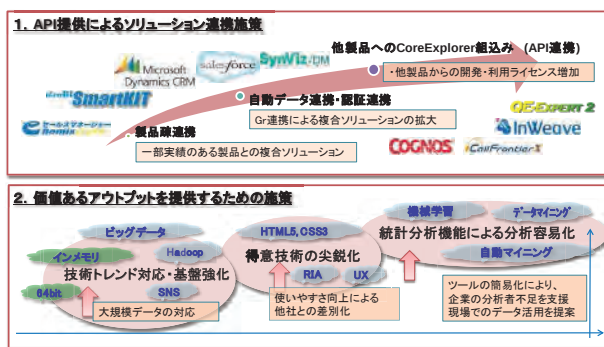


図4 更なる事業拡大に向けて

5. おわりに

昨今ソーシャルメディアやビッグデータの活用は、多くの企業にとって市場競争力を維持するための重要な課題となっている。また、データ活用の重要性が高まるとともに、今まで分析を専門的に行っていない部門・部署においてもデータを分析・活用することが求められている。実際にCoreExplorerを導入されたお客様の中にも、今までデータ分析を専門的に行っていないマネージャや設計担当者、営業担当者が活用している事例が増えてきている。

そのような背景の中、HSEでは11年間のCoreExplorerによるソリューションを展開してきた経験を基に、分析経験の浅いユーザでも結果を導きやすい機能や、分析作業を意識せずに他システムからテキストマイニング機能をシームレスに利用できるテキストマイニングツールを開発していく。そしてテキストデータ活用を企業に定着できるようにソリューションを展開し、事業の拡大を図っていききたいと考えている。

参考文献

- 1) 宮内秀彰他、日立 TO 技報第 18 号「テキストマイニングシステム CoreExplorer の今後の事業展開」
- 2) ITR, 「ITR Market View:DBMS/BI 市場 2014」



宮林 靖典 2003 年入社
KCMG
CoreExplorer, テキストマイニングツールの開発・適用
yasunori.miyahayashi.zu
@hitachi-solutions.com



宮内 秀彰 2008 年入社
KCMG
CoreExplorer, テキストマイニングツールの開発・適用
hideaki.miyauchi.zd
@hitachi-solutions.com



塚原 朋哉 1997 年入社
KCMG
CoreExplorer, テキストマイニングツールの開発・適用
tomoya.tsukahara.uh
@hitachi-solutions.com



川原 悠 2011 年入社
KCMG
CoreExplorer, テキストマイニングツールの開発・適用
yu.kawahara.hz
@hitachi-solutions.com



佐々木 智和 1999 年入社
KCMG
SynViz/DM, CoreExplorer の拡販・適用
tomokazu.sasaki.hz
@hitachi-solutions.com



佐藤 俊也 1993 年入社
KCMG
SynViz 製品, CoreExplorer の適用と
りまとめ
shunya.sato.ha
@hitachi-solutions.com