

テキストマイニングシステム CoreExplorer の今後の事業展開 ～事業実績と製品ロードマップ～

Future Business Deployment of the Text-Mining System CoreExplorer

(株)日立東日本ソリューションズ(日立 TO)ではこれまで、約 9 年間にわたりテキストマイニング製品である CoreExplorer の販売や、テキスト分析によるソリューションを展開してきた。製造業のお客様を中心に、公共・文教、流通・サービス、金融・保険、病院、研究所など様々な業種で分析の提案を行い、実績を積んできた。近年市場では、ソーシャルメディアやビッグデータの分析ニーズが高まっていることから、テキストマイニングの市場が拡大している。これら市場の動向やお客様からのフィードバックを踏まえ、CoreExplorer の事業計画の見直しを行った。今後の事業計画は、2012 年度下期で製品のメジャーバージョンアップを行い、事業拡大に向けた製品力強化を図る。その後の事業展開は CoreExplorer のクラウドサービスの展開、分析エンジンの API 化、さらに日立グループとの連携によりシナジー効果を生み出しいくことを目指している。

宮内 秀彰	Miyauchi Hideaki
川原 悠	Kawahara Yuu
塚原 朋哉	Tsukahara Tomoya
佐藤 俊也	Sato Shunya

1. はじめに

近年企業では、様々なテキストデータが電子化されており、これら蓄積されたデータを分析することによって業務の改善に役立てる様々な試みがされている。

製造業のお客様では、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、原因の早期解決や事故の防止に役立てている。またコールセンタに寄せられる顧客の声などからニーズの抽出や、クレームの要因分析に活用している。

これらの故障・不具合情報や、顧客の声は自由記述文で書かれることがほとんどであるため、大量のデータを人が一つ一つ読んで分析することは手間と時間がかかり、年間数万件のデータが蓄積される場合、すべてを読んで分析することは現実的に不可能である。

そういった課題に対し、日立 TO のテキストマイニングシステムである CoreExplorer を活用することで、短時間で定量的に分析を行うことを提案してきた。

本論文では、CoreExplorer を軸としたこれまでの事業実績を述べるとともに、お客様の評価や市場の動向を踏まえた今後の製品のエンハンス計画と、事業の展開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 これまでの取り組み

弊社のテキストマイニング事業は、2003 年にリリースした CoreExplorer Ver.1.0 から現在の Ver2.0 に至るまで約 9 年間継続して事業を行ってきた。製造業のお客様を中心とした品質分析や、お客様の声分析、また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を展開してきた。また日立グループ内の研究所や大学とも研究交流を図り、製品作りに生かしてきた。¹⁾²⁾³⁾

2.2 CoreExplorer の販売実績

2.2.1 導入実績(2003 年～2012 年)

CoreExplorer のこれまでの導入実績を表 1 に示す。製造業のお客様は 24 社(53%)と全体の半数以上を占めているが、その他の分野でも公共、流通、金融など、幅広い導入実績がある。これは CoreExplorer データ分析が、ある特定の業種に特化したものでなく、様々な分野で活用できることを意味している。

表 1 CoreExplorer Ver 2.0 の導入実績

#	分野	導入件数
1	製造業	24 社
2	公共・文教	8 社
3	流通・サービス	2 社
4	金融・保険	4 社
5	病院	3 社
6	研究所	4 社

2.2.2 導入顧客の評価

CoreExplorer の導入後、お客様から様々な評価や要望をいただいている。その中から主な評価内容を表 2、表 3 に示す。

機能面の評価点として、自動分類の精度が挙げられる。これは、あるルールにしたがって文書を自動的に分類する機能で、人が読まなくてもテキストデータの分類を行うことができる。⁴⁾ また、属性情報を用いた多角的な分析が行える点も評価されている。この多角的な分析を、単語のマウスクリックやプルダウンメニューの選択などの直感的な操作で行えるという機能は CoreExplorer の特徴として挙げられる。

一方、お客様からのご要望としては、一度の分析で取り扱えるデータ量を増やして欲しいといったことや、日本語だけでなく、英語、中国語といった言語の分析も行ってほしい、といった要望をいただいている。その他、製造業のお客様からは、統計的な分析や品質分析を行う上で必要な分析機能の充実も求められている。

これらお客様からの要望の多くは各業務での重要なニーズであることから、今後の製品開発の内容を検討する上で重要な要素として位置づけている。

表 2 CoreExplorer 導入の主な理由

#	顧客名	案件概要	評価点
1	製造業 T 社	製品の不具合分析	価格の安さ 自動分類の精度
2	製造業 F 社	営業情報・顧客の声分析	自動分類の精度 多角的な分析のしやすさ
3	製造業 J 社	製品の不具合分析	品質分析分野での実績
4	保険会社 K 社	FAQ 検索	価格の安さ 直感的な操作性

表 3 CoreExplorer への要望

#	顧客名	案件概要	ご要望
1	製造業 B 社	社内の技術情報、不具合データの分析	取り扱えるデータ量の 大規模化 品質分析に特化した 機能強化
2	製造業 T 社	不具合報告書の分析	英語、中国語へ対応
3	製造業 N 社	コールセンタの問い合わせ分析・FAQ 作成	分析機能の強化（統計分析など）

3. テキストマイニング市場の動向

近年のテキストマイニング市場のニーズとしては、アンケートやコールセンタに蓄積された顧客の声分析や、故障情報の分析のニーズが依然として高いが、これまでに比べ、分析データ量が増加傾向にある。また、国内企業が分析したい言語も中国語や英語といったニーズが増加している。その他、Twitter や Facebook などのソーシャルメディアの分析ニーズも高まっている。社内システム運用コスト低減の観点では、クラウド環境でのシステム利用ニーズも高まりつつある。

テキストマイニング市場規模は、2010 年度の国内市場の出荷金額が 24 億 6,000 万円で、前年比 11.3%増加しており、2011 年度以降もさらなる拡大が予想されている。⁵⁾ その背景として、ソーシャルメディアの情報を分析したいというニーズの高まりと予測されている。

これら市場の活性化を受け、シェアも変わりつつある。2007 年シェアトップである野村総研の TRUE TELLER のシェアは 60 %から 39%に減少し、プラスアルファコンサルティングのシェアは、前年比 40%増の 17.1%、IBM のシェアは 15.4%と大きく伸びている、大手企業がツールベンダを買収するなどの動きもあり、今後は様々なベンダの参入が考えられることから、競争の激化が予想される。

4. 現状の CoreExplorer の課題

現在の CoreExplorer での主な課題は、市場の動向やお客様のニーズから下記が挙げられる。

4.1. 大規模データに対する性能（数十万件規模）

近年の大規模コールセンタには 1 日数千件から数万件程度の問い合わせがあり、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、抽出したデータを分析する場合数百万件、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、数千万件から数億件のデ

ータ件数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化は重要な課題である。

4.2. 多言語の分析に未対応

近年日本の製造業は、海外に工場を持つことが多く蓄積される故障情報なども現地の言語で書かれている。また、販売先も海外の割合が増えてきていることから、多言語の解析技術が求められている。CoreExplorer は日本語解析専用のシステムであるが、分析言語の多言語化はグローバル企業をターゲットとする場合必要となる。

4.3. 品質分析に対する機能不足

日立 TO はこれまで、様々な製造業のお客様に対して製品開発や品質分析の支援を CoreExplorer で行ってきた。標準的な分析の手法である、FMEA や QFD の支援を行ってきたが、散布図による分散や相関の分析機能など、より定量的な分析を行う機能がユーザからは求められている。

4.4. クラウド化に伴うセキュリティの課題

システム運用コスト低減のため、クラウドサービスを導入する企業が増える中、CoreExplorer のクラウドサービスがあれば利用したいといったニーズもいただいている。CoreExplorer はスタンドアロン、もしくはプライベートクラウドでの利用形態であるため、パブリッククラウドへの対応も比較的容易に行える。しかしインターネット上に公開するため、不正アクセス防止など様々なセキュリティ対策が必要である。

4.5. CoreExplorer エンジンの部品 API 化

テキストデータの分析ニーズが高まっていることから、今後多くのソフトウェアでテキスト分析機能が求められると予想される。その際にテキストマイニングエンジンを API として提供可能であれば、様々な製品に組み込むことができるようになる。ライセンスビジネスによる事業の拡大を図るためには重要な課題の一つであるといえる。

5. CoreExplorer の製品エンハンス計画

前章で述べた課題を踏まえ、CoreExplorer は図 1 の製品ロードマップを予定している。

12 年度下期に現在の Ver.2.0 から Ver.3.0 へのメジャーバージョンアップを予定している。今後の事業拡大に向けた製品力強化と、製造業をターゲットとした品質分析ソリューションの強化を目的としている。Ver3.0 では下記の機能エンハンスを予定している。

5.1. スケーラビリティの確保

前章の課題 1 つ目でも挙げているように、顧客のニーズや市場のデータは大規模化してきている。CoreExplorer で扱えるデータ量はこれまで数十万件規模であったが、サーバのクラスタ化によってスケーラビリティを確保し、大量のデータにも柔軟に対応できる構成を実現する。これにより、大規模なコールセンタに寄せられる年間数百万件規模のデータ分析や、Twitter データの分析などが可能となる。

5.2. 中国語・英語の形態素解析機能実装

前章の課題 2 つ目に挙げている、分析言語に対する課題に対して、Ver3.0 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する。Mecab は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。そのため今後の言語拡張にも柔軟に対応することが可能である。本機能の実装により、英語、中国語の分析が可能となり、今後の市場の動向に合わせ適宜分析言語を追加することが可能となる。

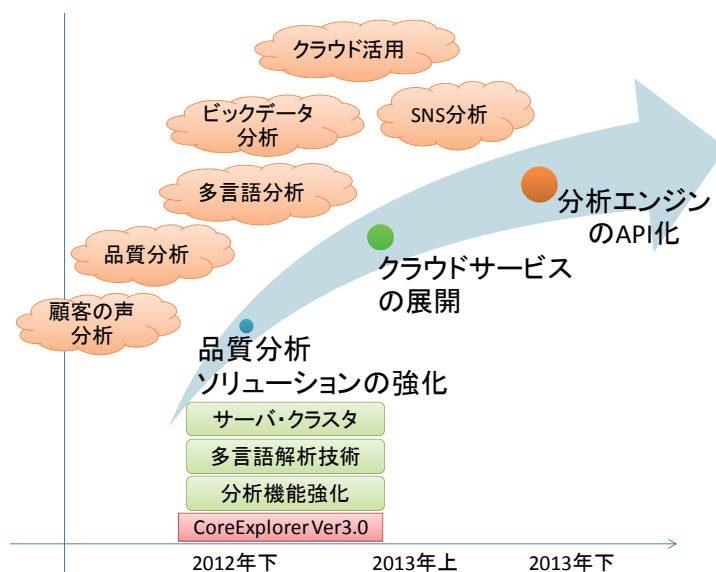


図 1 CoreExplorer の製品拡張計画

5.3. 品質分析機能の強化

前章の課題 3 つ目である、品質分析に対する機能の不足を補うために、品質管理の手法である QC7 つ道具の手法を実現する機能を実装する。具体的には、データのばらつきや相関を分析するための散布図や、散布図を各属性別に分析する機能を実装する。それにより、属性同士の関連性を分析することができ、ある現象に対する各属性の因果関係を知る手段として活用が可能となる。

これらの対応により、グローバル展開を図る国内の大手企業に対して積極的アプローチが可能となる。これまでの分析データ件数の不足や、分析言語の不足により、大手企業に対しての提案に課題があったが、Ver3.0 でこれらの課題が解決されることから、他社に負けない積極的な提案が可能となり、売り上げの拡大が期待できる。

6. 今後の事業展開について

日立 TO では、Ver3.0 のリリース以降も、さらなる事業拡大を目指している。事業拡大の要素としては、図 2 に示すように、大きく 2 つの事業内容を検討している。

一つは CoreExplorer のクラウドサービスの展開である。SNS から発信される大量なデータを各企業が蓄積し、運用するには大規模なサーバリソースを必要とする。そのためクラウド環境に蓄積したデータを CoreExplorer により SaaS 形態で提供することで、ユーザは自社に環境を構築することなく利用できるようになる。このような理由から、クラウドサービスを導入することで様々な業種、多くのユーザからの利用が見込める。

もう一つは、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。昨今のテキストマイニング市場の活性化から、様々な製品、システムでもテキスト分析のニーズは高まると予想される。

日立グループ内には、コールセンタシステムや、保全管理システム、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る。このように CoreExplorer を部品として利用することが可能となると、付加価値の高いソリューションを提供していくことが可能となり、グループ全体の売り上げ拡大に貢献できるものとして期待できる。



(C) 株式会社日立製作所, (C) 株式会社日立ソリューションズ
 (C) 株式会社日立ソリューションズ九州
 (C) 日立 INS ソフトウェア株式会社

図 2 今後の事業拡大の方針

7. おわりに

近年のテキストマイニング市場の活性化はソーシャルメディアによるものが大きいですが、それ以外に企業内のデータ活用のニーズも高まっている。弊社では約 9 年間、製造業のお客様を中心に多くのお客様データの分析を支援してきた事業実績があり、様々な業種でノウハウを蓄積してきた。

また、テキストマイニングによる分析のニーズは国内だけでなく、世界的にも広がりを見せると予想される。Facebook は全世界で約 9 億人、213 の国と地域で利用されている⁶⁾ (2012 年 3 月時点)。また中国のサービスである、微博(Weibo)は、約 3 億人(2012 年 3 月時点)を超える利用者がおり、日々多くのユーザの声が発信されている。

国内でのこれまでの実績を基に、海外市場に対してもテキスト分析のソリューションを展開し、事業の拡大を図っていきたいと考えている。

参考文献

- 1) 高梨勝敏他, 日立 TO 技報第 10 号「セマンティクス自動抽出によるエンタープライズオントロジの構築」
- 2) 塚原朋哉他, 日立 TO 技報第 13 号「複数視点からのテキストマイニングによる設計品質の向上」
- 3) 高梨勝敏他, 日立 TO 技報第 15 号「技術文書群からの技術・組織動向の抽出と可視化による戦略的な技術管理」
- 4) 渡邊まり子他, 日立 TO 技報第 17 号「東日本大震災復興に向けた支援活動～弊社製品を用いた被災者アンケートの分析～」
- 5) (株)ITR, 「ITR Market View : DBMS/BI 市場 2012」 2012 年 1 月 18 日
- 6) socialbakers, <http://www.socialbakers.com/>



宮内 秀彰 2008 年入社
ナレッジソリューショングループ
CoreExplorer, テキストマイニングツールの開発・適用
hideaki.miyauchi.01@hitachi-to.co.jp



川原 悠 2011 年入社
ナレッジソリューション G
CoreExplorer, テキストマイニングツールの開発・適用
munakata@hitachi-to.co.jp



塚原 朋哉 1997 年入社
ナレッジソリューショングループ
CoreExplorer, テキスト検索・分析システムの研究・開発
tomo@hitachi-to.co.jp



佐藤 俊也 1993 年入社
ナレッジソリューショングループ
CoreExplorer, テキスト・数値情報の分析ソリューションの適用
shu_sato@hitachi-to.co.jp