

# CoreExplorer Web API オプションのリリース と今後の事業展開

## CoreExplorer Web API options release and the future business development

(株)日立ソリューションズ東日本(以下、HSEと記す)は、テキストマイニングシステムであるCoreExplorer Ver4.0の提供を2016年5月に開始した。本バージョンでは、CoreExplorerの持つテキスト解析機能を外部システムから利用可能とするWeb APIオプションを追加した。Web APIオプションを利用することにより、自社の業務システムやパッケージ製品との連携が可能となり、テキストデータ活用機能を容易に組み込むことが可能となる。

本稿では、CoreExplorer Web API オプションの特徴や活用例の紹介と、今後の事業展開の方向性について述べる。

宮内 秀彰	Miyauchi Hideaki
塚原 朋哉	Tsukahara Tomoya
中野 修	Nakano Osamu
宮林 靖典	Miyahayashi Yasunori
原子 徹	Harako Toru
高橋 毅	Takahashi Tsuyoshi

### 1. はじめに

HSEは、テキストマイニングシステムであるCoreExplorerを2003年にリリースしてから、お客様の社内に蓄積された様々なテキストデータ活用のソリューションを展開してきた。

社内には、製品の製造過程で発生した故障情報や、メンテナンス情報、お客様からの苦情・クレーム、コールセンターの問い合わせログ、営業日報など様々なテキストデータが蓄積されているが、数値情報のように簡単に集計・傾向把握をすることは難しい。そこで、CoreExplorerを活用して、これらテキストデータを有効活用するための提案を、幅広い業種・部門で行ってきた。<sup>1)2)</sup>

提案の中では、CoreExplorerのテキスト解析機能を利用して、お客様の業務に合わせた個別システムを開発する案件も多く存在する。<sup>3)</sup>しかし、CoreExplorerを部品として利用するためには、インターフェースが未公開であることや、独自の仕様であることから、HSE内の一部の開発者しか設計を行うことができず、容易ではなかった。

また近年、「マイクロサービスアーキテクチャ」といった設計思想が注目を集めているように、「複数の特徴的な機能を持ったサービスを組み合わせることで、開発スピードを向上させたい」といったITベンダが増えており、CoreExplorerの引き合いでもテキスト解析機能を部品

として利用したいという要望が多くなっている。

そこで、CoreExplorerの持つテキスト解析機能を外部システムから利用可能とするWeb APIをオプションとして2016年5月から提供を開始した。

### 2. CoreExplorer Web API オプションの概要

#### 2.1 CoreExplorer Web APIオプションとは

CoreExplorer Web APIオプションは、RESTを採用することで容易にCoreExplorerのテキスト解析機能を利用できるオプションである(図1)。RESTとは、HTTPを利用しデータ操作をURLとパラメータで行うアーキテクチャを指す。GoogleやYahooなどが提供しているWeb APIもRESTで公開されており、Web APIの方式としては標準的な仕様である。CoreExplorerも本方式を採用し、導入時の学習コスト低減を狙っている。

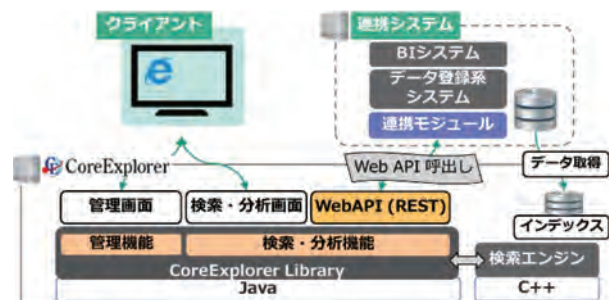


図1 CoreExplorer Web APIアーキテクチャ

## 2.2 Web API機能の特徴

Web APIはCoreExplorerの検索・分析画面で利用できる機能の他、今回新たに任意のテキストデータから単語を取得するリアルタイム単語取得機能や、文章(単語群)のカテゴリを推定するリアルタイム自動分類機能、急上昇単語だけを抽出する急上昇ワード抽出が利用可能である。Web APIで提供するテキスト解析機能を表1に示す。

表1 CoreExplorer Web API機能一覧

カテゴリ	機能名	概要
単語取得	登録文書単語取得	登録データの単語取得
	リアルタイム単語取得	任意のテキストから単語取得
文書検索	文書検索結果取得	検索条件にヒットした結果取得
	類似文書取得	単語の出現頻度が似たデータを取得
グループ表示	類似文書グループ表示	単語の出現頻度が似たデータをグループ化
分析	急上昇ワード抽出	指定した区間で急上昇した単語を取得
集計	属性×単語集計	指定した属性と単語のクロス集計結果を取得
自動分類	登録文書自動分類	登録データを分類ルールにしたがって分類
	リアルタイム自動分類	任意のテキストを分類ルールにしたがって分類

以下に各機能について、カテゴリ別に説明する。

### 2.2.1 単語取得機能

単語取得機能は、CoreExplorerに蓄積された大量のデータを解析し、単語一覧を取得する機能と、任意のテキストデータをリアルタイムに解析し単語を取得する機能の2つがある。

非構造のテキストデータを定量的な構造データに変換する基本的な機能で、単語への品詞タグ付けや、構文解析を行い、単語同士の係り受け関係解析結果を取得する。

### 2.2.2 文書検索機能

文書検索機能は、指定した検索条件にヒットしたデータを取得する検索と、指定した検索文字列に類似した文

書を検索する類似文書検索の2つがある。

利用ユーザが知りたい情報が明確な場合は、通常の検索機能を利用するが、蓄積されているデータがどのような単語で書かれているか分からない場合は、利用ユーザが知っているキーワードを検索条件として類似文書検索をすることで、得たい情報が検索しやすくなる。

### 2.2.3 グループ表示機能

グループ表示機能は、CoreExplorerに蓄積されたデータ内で、単語の出現頻度が似たデータをグループ化し出力する機能であり、文書クラスタリングとも呼ばれる。

似た文書をグループ化することで、データをすべて読まなくてもどういった話題の集合があるかを俯瞰することができる。

### 2.2.4 分析機能

分析機能の急上昇ワード抽出機能は、過去の期間（比較期間）に比べて最近（対象期間）の出現数が上昇している単語を抽出する。

急上昇ワードを日々監視することで、最近多発傾向にあるクレームの早期発見や、最近人気がある話題発見などに活用できる。

### 2.2.5 集計機能

集計機能の属性×単語集計機能は、指定した属性と単語のクロス集計結果を取得できる。

データ内に”年代”といった属性項目がある場合、年代毎にどのような単語が出現しているかを容易に比較することができる。

### 2.2.6 自動分類機能

自動分類機能は、CoreExplorerに蓄積された大量のデータに対して分類をする機能と、任意のテキストデータを自動分類する機能の2つがある。

どちらも自動分類するための分類条件を事前に定義し、分類条件に合致したデータに任意の分類名を付与することができる。

### 3. CoreExplorer Web API の活用例

#### 3.1 他システムとの連携イメージ

Web APIを利用することでCoreExplorerの検索・分析画面で表示している検索・分析結果を連携対象のシステムで利用することが可能となる。

Web APIを利用した他システムとの連携イメージを図2に示す。

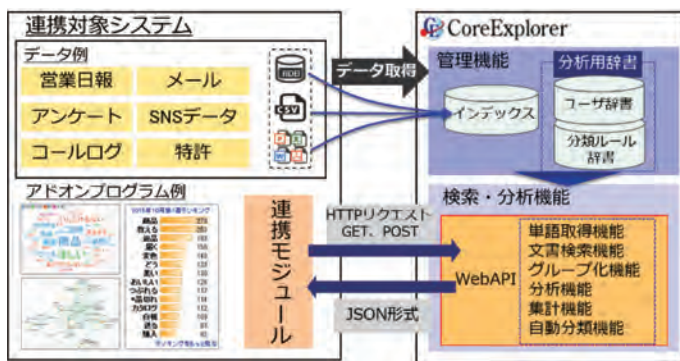


図2 CoreExplorer Web APIによる連携イメージ

データ連携は、連携対象システムで保持している営業日報やアンケートなどのテキストデータをCoreExplorer管理機能でCoreExplorerに取り込む。データ取り込みはCoreExplorer標準機能でCSVデータやRDBデータの自動取り込みが可能である。

取り込んだデータは連携モジュールを通して検索・分析データとして連携システム側で取得し、ポータル画面やアラートメール配信に利用することが可能である。

現在、CoreExplorerのWeb APIを活用して様々なパターンで提案活動を行っている。その中から、他システムとの連携による活用例を3つ紹介する。

表2 他システム連携による活用例

連携対象システム	期待する効果
BIツール	定性データ(自由記述文)の情報もBIツールのレポート対象とできる。
音声テキスト化システム	音声コールログから、問い合わせ概要を把握。FAQ拡充。
ソフトウェア品質管理システム	情報検索の効率化。品質向上観点の抽出。

#### 3.2 活用例① BIツールとの連携

分析に活用できるデータは大きく分けると定量データ(数値情報など)と定性データ(自由記述文など)がある。例えば売上の推移を分析する場合、通常は数値情報をBIツールに取り込み、月別日別などで分析を行っている。しかし、数値情報だけでは増加減少の結果はわかるが、なぜ減少したのかの理由まではわからない。そのため減少の理由を知るために要因の一つとしてクレームの情報や問い合わせ情報などを活用したいという要望がある。

BIツールとテキストマイニングツールはそれぞれが独立したツールであることが多いため、それぞれのツールを使い出力した結果から人手で関係性を分析する手法がほとんどである。しかし、それぞれのツールには関連性がないため、より効率的に定量データと定性データを組み合わせて新たな価値を発見することは困難である。

CoreExplorerのWeb APIでは、定性データ(自由記述文)からキーワードのランキングや自動分類結果などの分析結果を出力できる。出力データをBIツールで取り込むことのできる形式で提供することで、テキストマイニングの結果と定量データを組み合わせた集計、分析が可能になる。これによりこれまで以上にデータ活用が向上する。連携イメージを図3に示す。

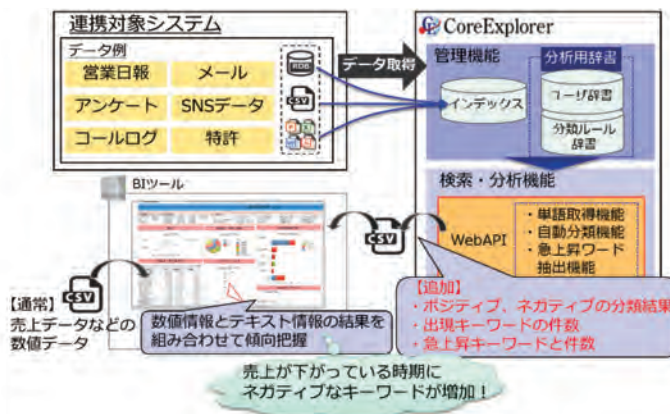


図3 BI ツールシステム連携概要

定量データと定性データを組み合わせた活用としては以下のようなことが考えられる。

- 1) 売上の情報とクレーム情報の関係性を分析  
売上増減の要因情報の一つとしてクレームに含まれるキーワードの変化やポジティブ、ネガティブなどの評判と数値情報を組み合わせた分析を実施。

2) 故障情報件数と保守点検記録の関係性を分析

設備の故障が発生した前後の保守点検記録ではどのようなキーワードを含んだ報告が増加していたかの関係性を分析。

3.3 活用例② 音声テキスト化システムとの連携

顧客接点を担っているコールセンタには「顧客の声」が寄せられ、その「声」をサービスの品質向上や新製品開発など企業活動に活用しようと考えている企業は多い。

昨今は音声テキスト化システムを用いて、「声」の音声データからテキストデータへと、分析しやすい形式に変換して扱うケースが増えている。

日立製作所でも、「顧客の声」を分析し、企業の課題解決を支援する音声データ活用ソリューションを展開しており、CoreExplorer は本ソリューションで扱う分析ツールの一つとして位置づけられている。

音声のテキストデータを活用することにより、オペレータの回答履歴入力業務の効率化を図りつつ、会話の中に隠された有益な情報や、問題の兆候を見つけ出すなど企業の期待は大きい。その一方で、会話の内容がすべて記録される音声のテキストデータは膨大になり、目的に沿ったデータを抽出するための作業負担が高く、工夫が必要である。

そこで CoreExplorer Web API を活用することにより、予め着目すべきデータを抽出・可視化することで、効率的なデータの利用が可能となり、企業の課題解決を支援できると考えている。活用例を以下に2つ示す。

1) 特徴的なキーワードによる問い合わせ傾向の把握

単語取得機能や急上昇ワード抽出機能により、問い合わせに頻出する特徴語や多発傾向にあるワードを抽出し、属性の集計では把握できない問い合わせ内容の傾向を把握する。システム連携イメージを、図4に示す。

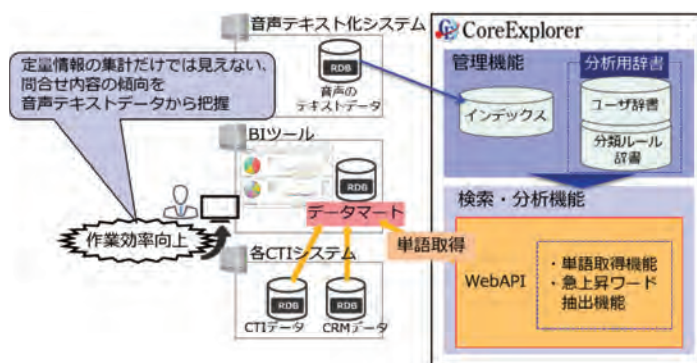


図4 音声テキスト化システム連携概要1

2) サービス品質向上に向けた、FAQの作成候補の抽出

表1記載の、類似文書取得機能により、作成済みのFAQと比較し、既存FAQにない「顧客の声」を抽出し、さらにグループ表示(文書)機能を用いてFAQ作成候補としてまとめる。システム連携イメージを図5に示す。

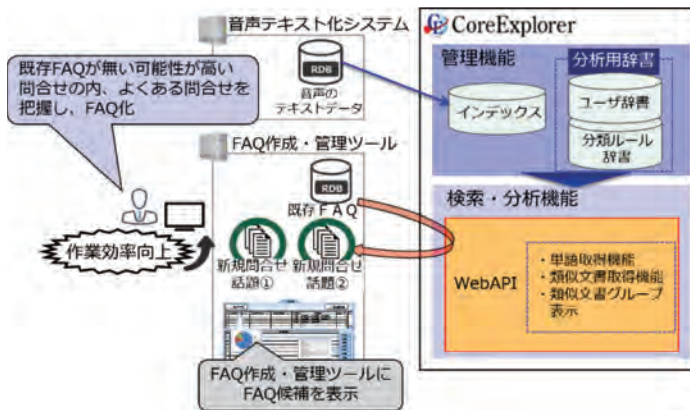


図5 音声テキスト化システム連携概要2

3.4 活用例③ ソフトウェア品質管理システムとの連携

ソフトウェア開発のテスト工程では、障害の発生工程や画面名、機能名などの属性情報と、障害の詳細状況、原因などの自由記述情報を不具合管理票として記録し、品質管理システムで管理している。管理されたデータの属性情報は、集計することで定量化しソフトウェアの品質評価に利用している。

一方、自由記述情報は1件ずつ読む必要があり、対象となる障害1件を修正する用途だけで利用され、他のデータとの関連や品質評価では、活用はされてこなかった。

そこで、CoreExplorer Web API を活用することにより、ソフトウェア障害の自由記述情報に対して新たな観点での活用が可能となると考えている。活用例を以下に2つ示す。

1) 過去データの類似検索による原因修正工数の短縮

類似文書取得機能により、現在発生している不具合情報から、過去の類似不具合管理票を検索・提示することで、原因調査に掛かる工数の短縮を図る。システム連携イメージを、図6に示す。

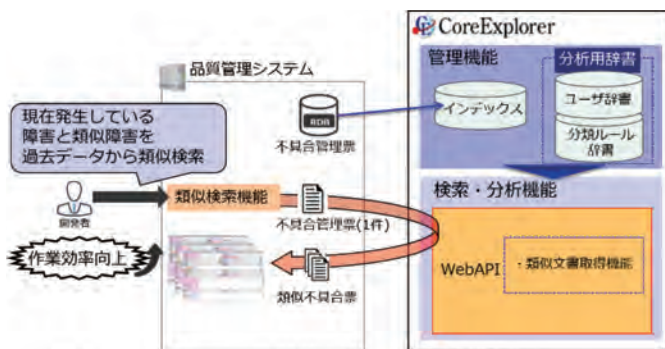


図 6 ソフトウェア品質管理システム連携概要 1

2) 残存不良観点の提示による品質向上

表 1 記載の属性×キーワードのクロス集計機能により、各テスト工程に出現しているキーワードを比較することで、どの工程でも出現するキーワードは修正が不十分な可能性があるとして判断できる。そこに残存不良の可能性があると想定し、テスト項目を追加し、品質向上を図る。システム連携イメージを図 7 に示す。

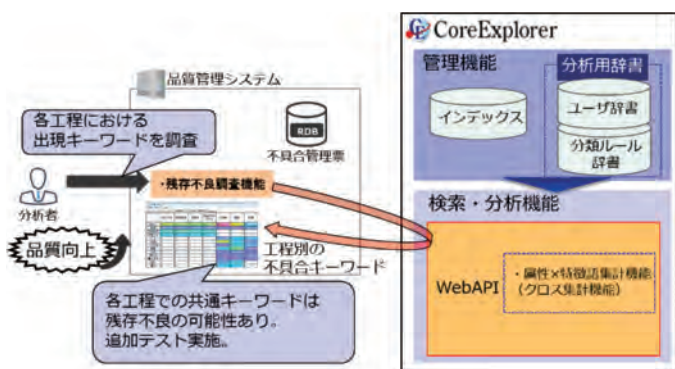


図 7 ソフトウェア品質管理システム連携概要 2

4. 今後の事業展開

4.1 テキストマイニング市場の動向

国内でのテキストマイニング市場は、2016 年度は前年比 109.5%の成長で推移しており、2016 年度は 46 億円程度の規模が見込まれている。2017 年以降も拡大基調にあると予想されており、今後も成長が期待される分野であることが分かる。<sup>4)</sup>

近年では、マシン処理能力の向上やそれに伴い複雑な処理が可能になったことで自動運転や人工知能などの自律的なコンピュータが台頭しており、テキストマイニングだけでなく、機械学習や音声認識技術などのコア技術を複合して新しい価値の提供が期待されている。

また、冒頭でも述べたように、「マイクロサービスアーキテクチャ」の設計思想に基づいたシステム開発が一般的になることで、テキスト解析 API の需要も高まると考えられる。

4.2 HSE内製品連携での今後の事業展開

HSE における今後の事業展開の主なターゲットとしては、2016 年 4 月にリリースした「業務アプリ構築プラットフォーム AppSQUARE(アップスクエア)」との連携があげられる。AppSQUARE は、ノンプログラミングで画面項目を定義できるフォーム機能とプロセス管理機能を備えており、社内のデータを一元管理するために優れた製品である。

この AppSQUARE に、CoreExplorer の Web API を連携することで、AppSQUARE のオプションとして高度な検索機能を追加できる。

例えば、検索キーワードをリアルタイム単語取得により分解し、分解された単語で部分一致検索を行うことで、検索結果の網羅性を向上させることができる。

AppSQUARE による社内のデータ管理と、蓄積されたテキストデータの活用をトータルで提供することによって、他社との差別化を図っていきたいと考えている。

4.3 日立グループとの連携強化による事業展開

日立グループとの連携も活発化してきている。日立製作所は、「顧客の声」を分析し、企業の課題解決を支援する「音声データ利活用ソリューション」の販売を 2016 年 3 月に開始した<sup>5)</sup>。音声データのテキスト化には、Deep Neural Network 技術型の音声認識エンジンを持つ「Speech Recognition Platform」を採用することによって、高い認識率を実現している。テキスト化されたデータであれば CoreExplorer での解析が可能となるため、音声データの活用としても CoreExplorer が期待されている。

4.4 今後の製品エンハンスについて

CoreExplorer Ver4.0 では、これまでの CoreExplorer が持つテキスト解析機能を外部から利用できるように Web API をオプションとして追加し、様々なシステムと連携しテキストデータの活用が提案可能となった。今後は、テキスト解析基盤として広く利用していただけるように、以下のような機能拡張を検討している。

① Web API による提供機能の拡充

CoreExplorer の Web API は、取得した単語の集計結果やキーワード一致による分類などの基本的な情報取得と、単語同士や他の属性項目との関連から新たな知見を得る、高度な情報取得に大別される。

基本的な情報だけでは大量の単語から知見を得るのは難しいため、多変量解析などによる高度な情報取得の機能拡充をすることによって、業務でのテキストデータ活用をしやすいことを図っていく。

② 自動分類ルール定義の簡易化と分類の高精度化

データの自動分類は、テキストデータ活用の良くあるニーズの一つである。現在の CoreExplorer では、自動分類するためのキーワードを事前に定義しておくことで、データを自動分類することが可能である。しかし分類ルール定義は人手を要するため、分類ルール定義の簡易化が課題となっている。また業務活用時のポイントとしては、自動分類の高精度化も重要な課題であるため、今後の機能拡張に力を入れていきたいと考えている。

③ 単語へのメタデータ付与

現在の CoreExplorer では、形態素解析によって、日本語文書の分かち書きと品詞付与は実現している。しかし分かち書きされた単語には、品詞以外にも、製品名や素材名、ポジティブな意味の単語、ネガティブな意味の単語など、業務的な意味を多く含んでいる。大量な単語情報から上記のようなメタデータでフィルタリングをすることによって、よりノイズの少ない情報取得可能となると考えられる。

5. おわりに

昨今、社内データの利活用だけでなく、社外のオープンデータの利活用が官民共同で推進されている<sup>6)</sup>。その中には、数値情報以外の非構造データも多く含まれており、テキストデータもその一つである。これらテキストデータを扱うためにも、テキストマイニングはコア技術の一つとして必要となる。

CoreExplorer Ver4.0 は、Web API オプションによって、様々なシステムと簡易に連携できる仕組みを提供することが可能となった。Web API によって実現できる様々な活用例を紹介したが、社内外に蓄積されているテキストデータの種別は多岐にわたるため、様々な活用例が考えられる。

CoreExplorer は今後テキストデータ活用基盤としての機能拡張を図り、多くのシステム・サービスの基盤技術として、社会イノベーションに貢献していく所存である。

参考文献

- 1) 宮林靖典他, HSE 技報第 20 号「CoreExplorer Ver.3.0 のリリースと今後の事業展開」
- 2) 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」
- 3) 宮内秀彰他, 日立 TO 技報第 15 号「お客様の声分析システム VOCAnalyzer の開発」
- 4) ITR, 「ITR MARKET VIEW DBMS/BI 市場 2015」
- 5) 日立製作所, ニュースリリース「音声データ利活用ソリューション」  
<http://www.hitachi.co.jp/New/cnews/month/2016/03/0323.html> (2016 年 9 月閲覧)
- 6) 総務省, ICT 成長戦略「オープンデータ戦略の推進」  
[http://www.soumu.go.jp/menu\\_seisaku/ictseisaku/ictriyou/opendata/](http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/) (2016 年 9 月閲覧)



宮内 秀彰 2008 年入社  
KCG  
CoreExplorer の事業推進  
hideaki.miyauchi.zd@hitachi-solutions.com



塚原 朋哉 1997 年入社  
開 11G  
CoreExplorer の製品開発  
tomoya.tsukahara.uh@hitachi-solutions.com



中野 修 1998 年入社  
KCG  
AppSQUARE, CoreExplorer の提案・構築  
osamu.nakano.dc@hitachi-solutions.com



宮林 靖典 2003 年入社  
V2G  
CoreExplorer の提案・構築  
yasunori.miyahayashi.zu@hitachi-solutions.com



原子 徹 2006 年入社  
S 営 2G  
CoreExplorer の拡販  
toru.harako.kk@hitachi-solutions.com



高橋 毅 2013 年入社  
KCG  
CoreExplorer の提案・構築  
tsuyoshi.takahashi.rd@hitachi-solutions.com