

表1 主要なテキスト・マイニング・ソフトウェアの一覧（国内）

No.	製品・サービス名	開発元・販売元	特徴	守備範囲
	g Server ソフトウェア	富士通(株)	キーワード間の関連性をビジュアルに表示する「コンセプトマップ」。OLAP 製品と組み合わせ使用可能。 http://software.fujitsu.com/jp/symfoware/products/textmining/	
		(株)日立製作所	ターム（単語・語句）の共起関係による相関分析・分類、自然文検索、概念検索など、統合的文書管理システム http://www.hitachi.co.jp/Prod/comp/soft1/docbro/	
3	TAKMI マイニングシステム	日本アイ・ビー・エム(株)	概念（キーワードとなる文字列とそのカテゴリー）を抽出し、定型情報と共に統計量を計算・結果表示。 http://www.trl.ibm.com/projects/s7710/tm/takmi/takmi.htm	
	le Meister マネジメントシステム	(株)東芝	キーワードの出現頻度・関連度によるクラスタリング、依存・品詞分析によるテキスト・マイニング（要因分析） http://cn.toshiba.co.jp/prod/km2/function/mining.htm	
5	Knowledgeocean(ナレッジオーシャン) ナレッジマイニング支援システム	(株)NTT データナレッジ	コンセプト（主要語、概念）の抽出によるコンセプトの共起分析、クラスタリング、類似文書検索 http://www.knowhowbank.com/html/sol/sol_kso_1.html	
6	MiningPro21 文書マイニングシステム	日本ユニシス(株)	単語の相関度による文書分類、連語抽出・判別関数による文書判別、日本語文章による類似文書検索 http://www.unisys.co.jp/MP21/bunsho/	
	CB Market Intelligence テキストマイニング・ソリューション	(株)ジャストシステム	意味認識手法（自然言語処理技術がベースのテキスト分析技術）による主題・評価・感性・機能要求分析 http://www.justsystem.co.jp/cbmi/	
	専用ツール テキストマイニングツール	クオリカ(株) (旧コマツソフト)	コンテキストベクタ（似た文脈の中で用いられる単語のベクトルは似た方向を持つ）方式による知識モデル生成 http://www.qualica.co.jp/develop/vextminer/	
9	DE-FACTO	電通，電通リサーチ	発想支援ソフト，テキスト型データから単語・語句の関連性を重要度に応じて類型化し，視覚化する．	
10	SurveyAnalyzer(サーベイアナライザー) 自由記述アンケート分析システム [TopicScope として改編されたもよう]	日本電気(株)	確率的コンプレキシティ（統計尺度）に基づき、分析対象と結びつく固有の言葉や語句を抽出・発見 http://www.nec.co.jp/press/ja/0110/0502.html	
11	Text Mining for Clementine (LexiQuest) 調査データ分析 専用ツール	エス・ピー・エス・エス(株)	コンセプト（意味ある言葉の組み合わせ）の抽出。データ・マイニングツール Clementine のプラグインツール http://www.spss.co.jp/product/cle_text/text.html	
12	自由回答設問他 統計的手法 レーテラー) マイニング分析システム	(株)野村総合研究所	係り受け(主語・述語)構文解析、話題・因果関係マッピング、文書スコアリング、分析結果の EXCEL 出力 http://www.trueteller.net/	
13	WordMiner(ワードマイナー) テキスト型データ解析ソフトウェア	日本電子計算(株)	構成要素（単語や語句）抽出による多次元データ解析（対応分析、クラスタ化）、コンコーダンス（用語検索） http://www.jip.co.jp/bs/products/Shohin/52/kihon/kihon.html	

会社名、製品名等は、各社の登録商標もしくは商標

表2 欧米のテキスト・マイニング・ソフトウェアの例

No.	製品・サービス名	開発元・販売元	特 徴
1	Sphinx Survey Plus2 & Lexica	Le Sphinx Développement SCOLARI http://www.scolari.com http://www.pugh.co.uk/Products/scolari/surveyplus.htm http://www.lesphinx-developpement.fr/	<ul style="list-style-type: none"> ・調査データの集計・分析を主とする ・内容分析, 文脈分析を行う ・多変量解析 (主成分分析・対応分析など)
2	SPAD.T (Système Portable pour l'Analyse des Données-Donnée Textuelles)	L. Lebart (ENST) とそのグループ	<ul style="list-style-type: none"> ・記述的・探索的ツール ・調査データ (自由回答など) の解析を重視 ・選択肢型設問とのクロス分析 ・多変量解析 (対応分析, クラスタ化) ・単語・語句の有意性テストによる特徴抽出 ・コンコーダンスによる単語・語句の利用パターン観察 ・WordMiner の元となったソフト
3	WORDSTAT (V4.0)	Provalis Research Inc. http://www.simstat.com/home.html	<ul style="list-style-type: none"> ・内容分析を主とする ・統計ソフト SIMSTAT , CodeMiner にリンク (*)CodeMiner: Qualitative Data Analysis Tool
4	STATISTICA Text Miner	StatSoft Inc. http://www.StatSoft.com/ http://www.StatSoft.com/textminer.html	<ul style="list-style-type: none"> ・統計ソフト STATISTICA と併用 (add-on), 統計処理機能の利用 (PCA, k-means クラスタ化, その他のデータマイニング) ・STATISTICA に渡す前の事前処理 ・種々のテキスト・フォーマットに対応 ・削除機能とそのルール, stub-list の生成 ・stemming algorithm の適用 ・多言語対応 (オランダ, ドイツ, 英語, フランス, イタリア, ポルトガル, スペイン, スウェーデンなど) ・文章要約化の機能 ・SVD (特異値分解) による特徴抽出
5	Text Analysis	MEGAPUTER Inc. http://www.megaputer.com/	<ul style="list-style-type: none"> ・セマンティック・テキスト・マイニング: キー概念と非構造的テキスト型ノードとの関係から意味論的 (セマンティック) 分析を行う ・Link Analysis を使って, 意思決定に役立つような視覚化を行う
6	WEBSOM	Helsinki University of Technology http://websom.hut.fi/websom/	<ul style="list-style-type: none"> ・ドキュメント探査ツール, 視覚化ツール ・Self-Organizing Maps (SOM) を使う ・Kohonen が主催するグループの研究公開

表3 テキスト・マイニングの位置づけ

データの型 (種類)		対 象	目 標 と 対 応		
			対応方法 適用の方法論	単純なパターンの発見	意味ある複雑な情報の発見
数値型データ	質的データ (名義, 順序) 量的データ (区間, 比例)	・テキスト型データを計量化・数量化し, 数値型データとみなして処理 ・数値型データとテキスト型データの併用	典型的なデータマイニング 統計解析手法 (*) 特徴, 傾向, 規則性の探査・発見 (*) モデリングの支援	・データベース問い合わせ (*) 単純な検索, 情報アクセス, 参照など ・タグ化, コード化, カテゴリー化など ・情報検索, 情報抽出	・形式的 TM の実行 (*) テキストの計量化を通じて探査 (*) 現状の TM の主流 これで「意味ある複雑な情報の発見」は本当に可能か
		小説, 自由記述文, 自由回答など (非構造的)	自然言語処理・計算機言語学 (*) 構文解析, 意味解析, 文脈解析, ... (*) 共起, 係り受け等	文書要約 文書分類 内容分析 全文検索	真の TM とは? さらに有効な TM の方法はあるのだろうか?
一般文書類 (構造的)					
非数値型データ	非テキスト型データ	画像, 音声など	計算機言語学 言語学 音声学	自動翻訳技術 多言語間翻訳	自動通訳機能 意味理解 音声認識 画像認識

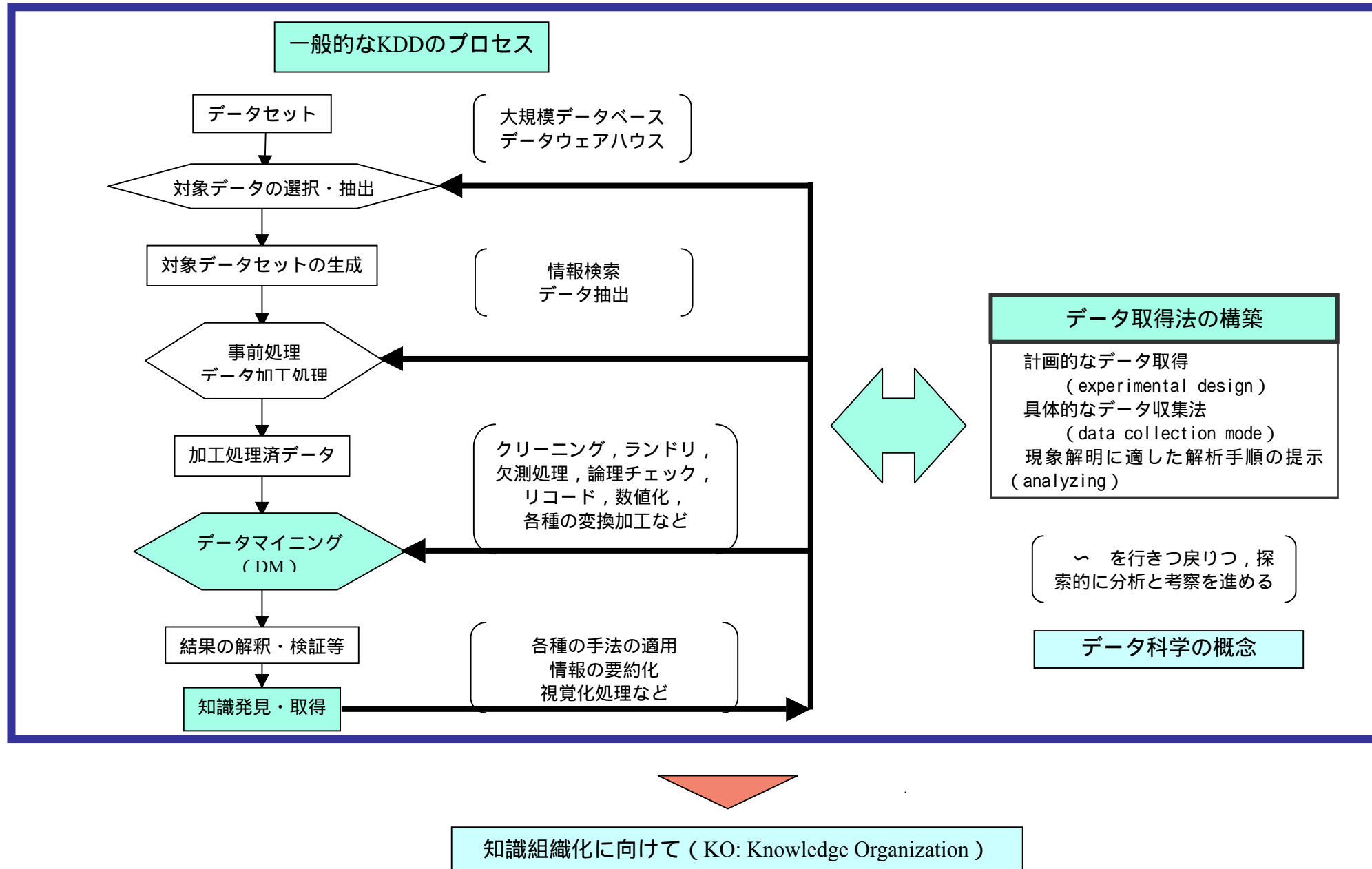


図1 KDDにおけるDMの位置づけとデータ科学の関係



図2 テキスト・マイニングに関連する研究分野



図3 テキスト・マイニングに利用される主な方法論

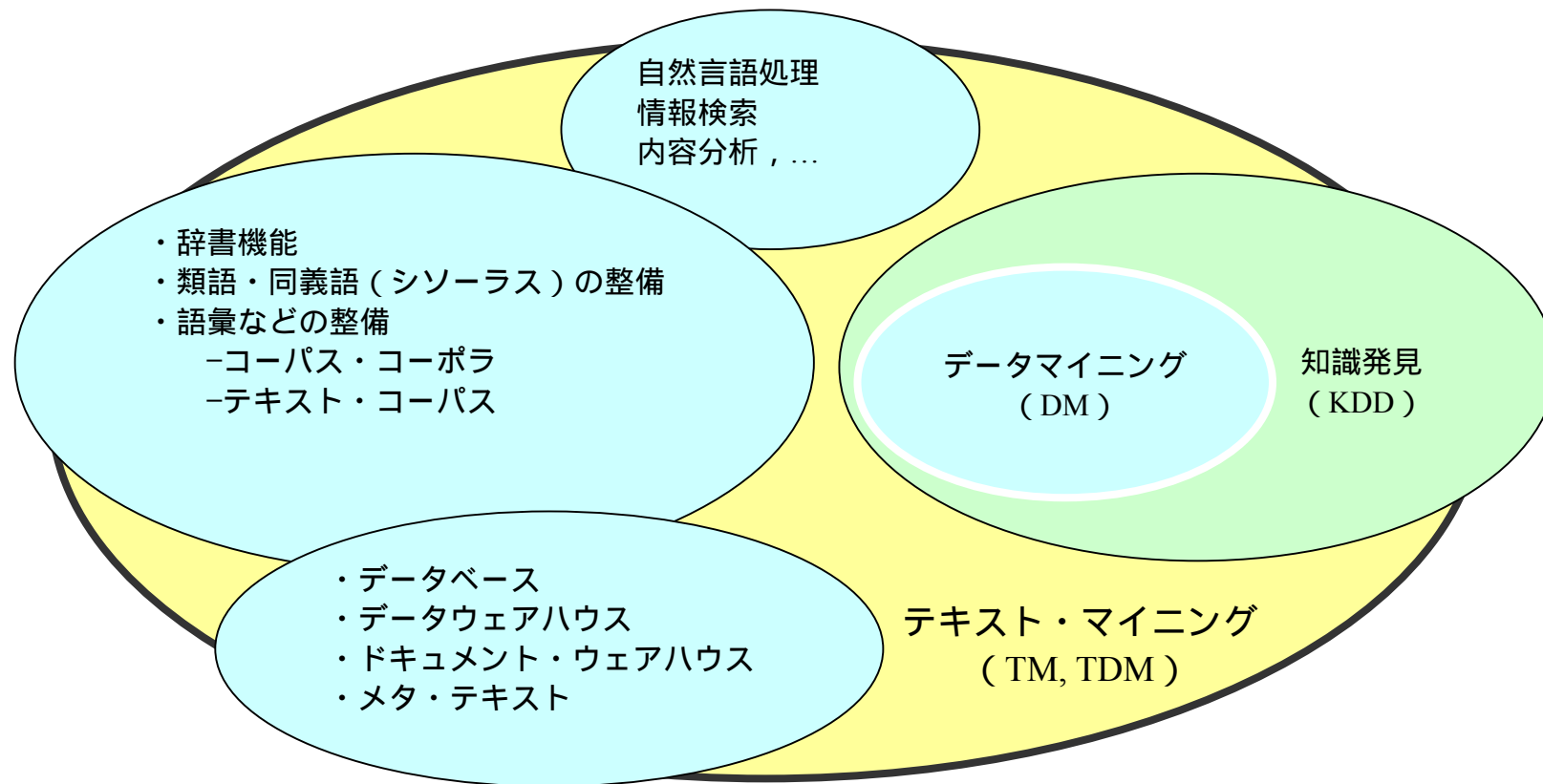


図4 KDD, DM, TM の関係