

eマーケティングとText Mining Studio

- 大学教育および研究ツールとしての
Text Mining Studio活用事例 -



立教大学 社会学部
教授 佐々木 宏

報告内容

コンテキスト・マイニング

テキストマイニングの一般ステップと留意点

TMS 活用事例

(大学教育および研究ツールとして)

以下、Text Mining StudioをTMSと表示



コンテキスト・マイニング

eマーケティング

キーワード: コンテキスト発掘

eマーケティング

- eビジネスのマーケティング戦略
- 一般企業のWebを活用したマーケティング戦略

○ eマーケティング実践のための基本情報ニーズ

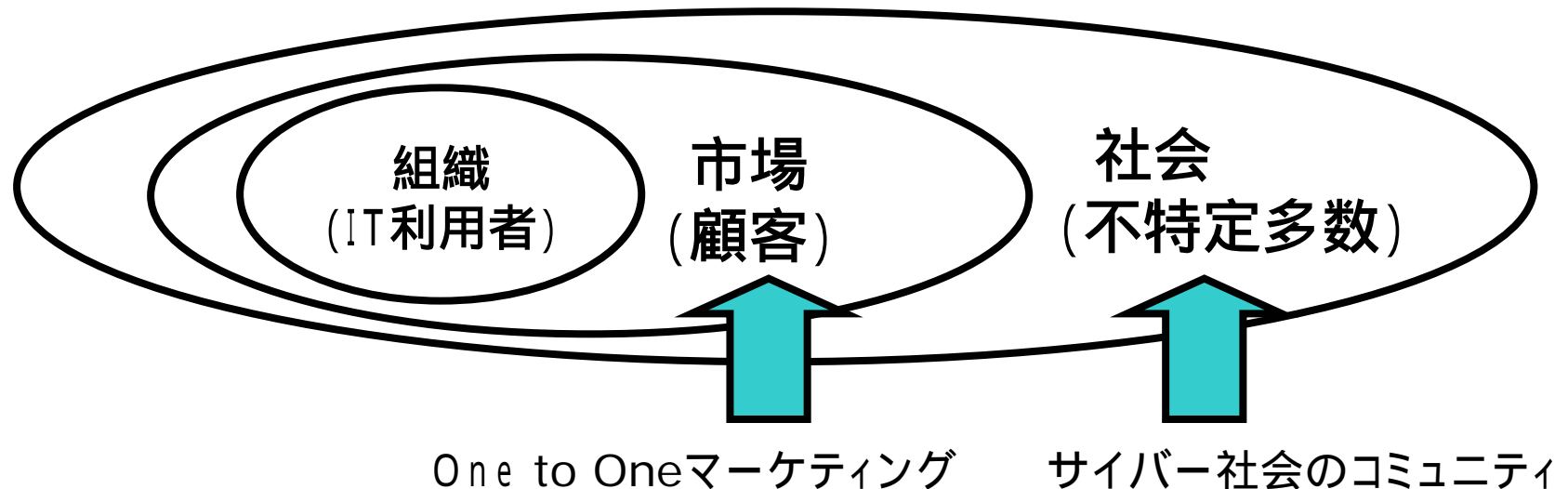
- 何が売れているのかではなく、どのようなコンテキストで売れているのか？
- デジタル・コンテンツを活用し、そこから事業戦略上意味あるコンテキストを発見したり、新しいコンテキストを創造していく
- その構造や変化をいち早く可視化するところに、基本ニーズがある

○ 情報の遍在から偏在へ

- サイバー社会に遍在し、膨張を続けるデジタルデータをどのように構造化するか？
- スピードとダイナミックな変化

発掘すべきコンテキストの所在

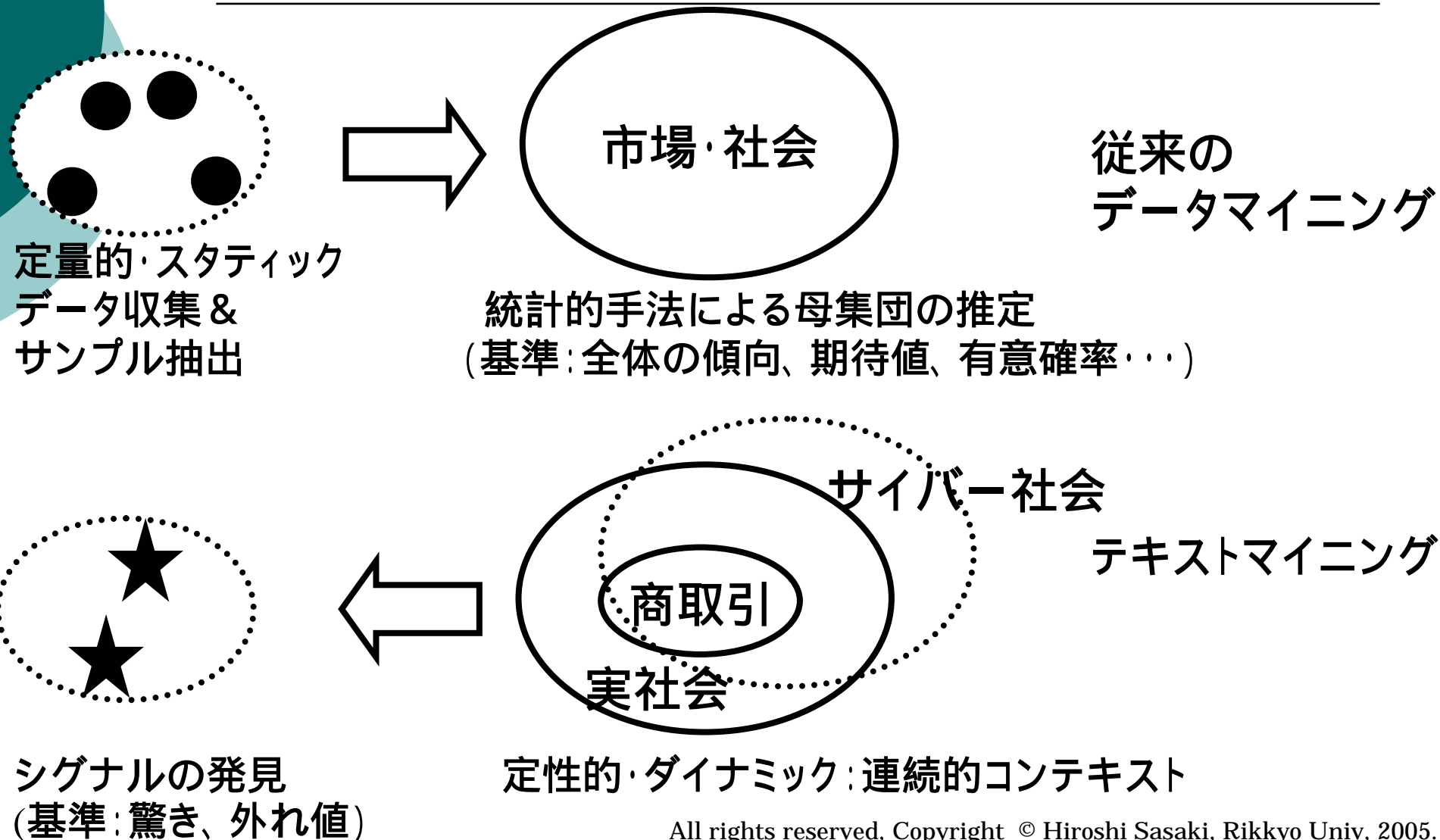
- コンテキストの所在
 - どの、どのデジタルデータをソースにするのか
 - 3つの対象領域からコンテキストを発掘、発見する



コンテキスト・マイニングとは？

- 情報の2面性
 - コンテンツとしての情報
 - デジタル情報:サイバー社会のコンテンツ
 - コンテキストとしての情報
 - コンテキスト:前後関係、文脈、背景、状況などの意味
 - 情報はコンテキストを形成
- コンテキスト・マイニング
 - デジタルデータというコンテンツを活用して市場ないしは社会のコンテキストを発見する手続き
 - テキストマイニングはそのためのツールのひとつ
- コンテキスト・マイニングのための2種類の情報ソース
 - トランザクション :商取引(エコノミック関係)でクローズ
 - コミュニケーション :社会(ソシオ関係)にオープン
- BI (Business Intelligence) 構築のためのデータ
 - これまでの定量データから、定性データの取り込みへ

コンテキスト・マイニングにおける テキストマイニングの位置づけ

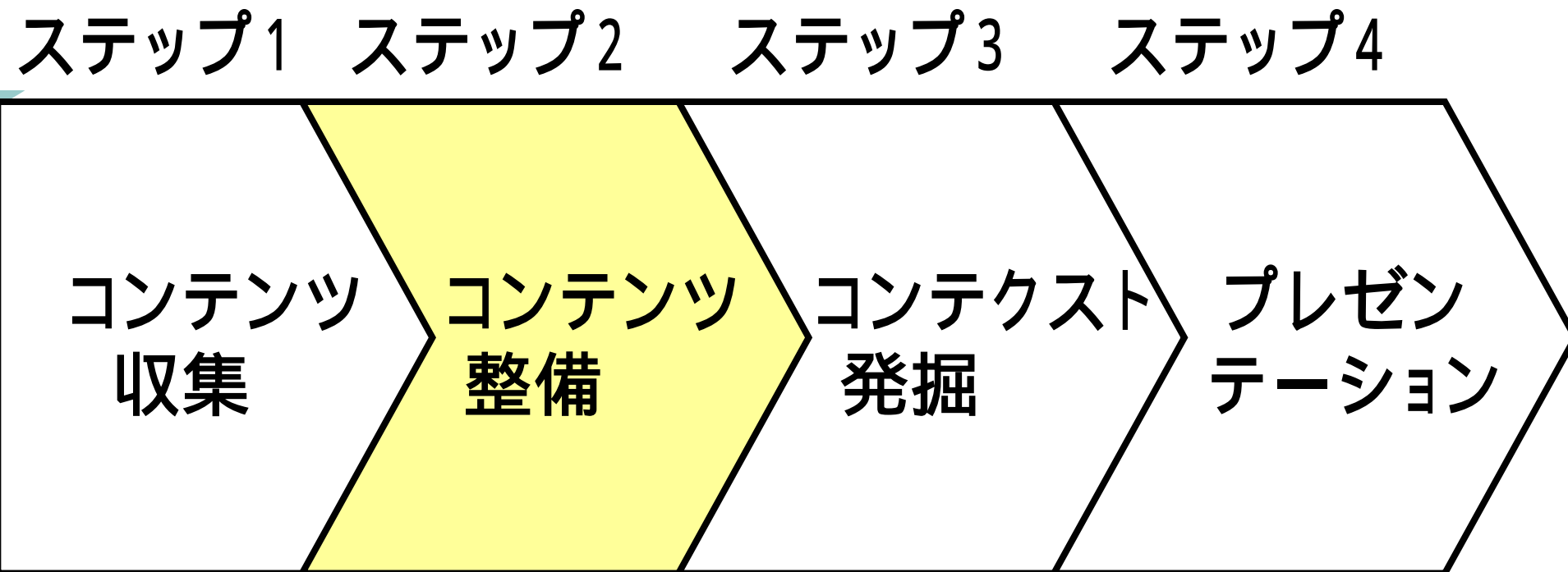




テキストマイニングの 一般ステップと留意点

テキストマイニングの基本ステップ

以下は、テキストマイニングの基本ステップを一般化したものである。各ステップの留意点、問題点にはどのようなものがあるか？



テキストマイニングのプロセス

ステップ1 コンテンツ収集

- あらゆるところに遍在するコミュニケーションデータ
 - 商取引に付随するコミュニケーションやサイバー社会のオープンなコンテンツは鮮度の高い宝の山であるという認識
 - 対象をどこに定めるか(コンテンツの所在)
 - 顧客とのコミュニケーション
 - 商談やアフターサービスなど
 - Web上のオープンなコミュニケーション
 - 市場に製品を投入すれば、たちまち多くの風評がWeb上に流れ始め、それが自己組織化し、さまざまなコンテキストを形成していく。
- マーケティングデータ: アンケートなどの自由記述欄

テキストマイニングのプロセス

ステップ2 コンテンツ整備

メッシーなコンテンツをソフトウェアで分析できる形に整備しなければならない。

このステップに最も手間がかかる。

データ構造の整備 (次ページ)

文章(日本語)の整備:形態要素解析ソフトを意識

- 文章内の単語同士の関連(主語述語の係り受けなど)
- ことばの意味(『やばい』、『大丈夫です』の意味は?)
- 述語属性
 - 否定、可能/不可能、要望、疑問、困難 (その他 理由)
 - 日本語独特の微妙な言い回し(「よい」「良い」「いい」「いいかもしれない」「よくないかもしれない」「悪い」「良くない」「良くなくない」…など)
- 辞書:分析目的に応じたテクニカルタームなどの抽出

テキストマイニングのプロセス

ステップ2 コンテンツ整備

データ構造の整備 (前ページ)

トランザクションデータ (多次元構造)

日時	店舗	顧客	・・・	販売高
----	----	----	-----	-----

商取引に付随するコミュニケーションデータ

日時	店舗	顧客	・・・	テキスト
----	----	----	-----	------

サイバー社会のオープンソース

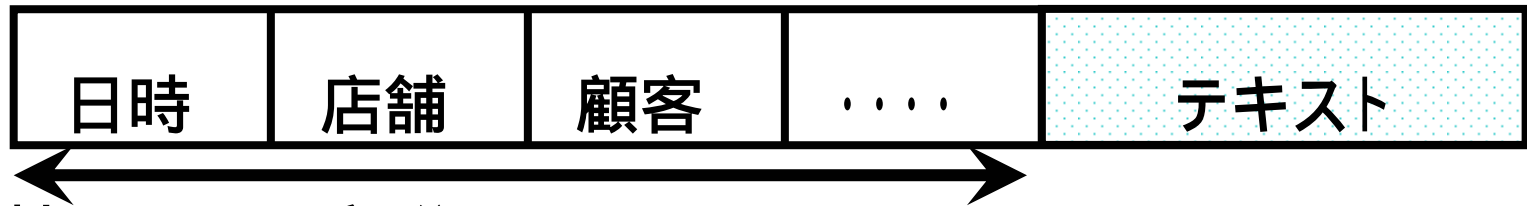
不定形



多次元構造化

テキストマイニングのプロセス

ステップ3 コンテキスト発掘



○ 属性とテキスト部分との関わり

- 従来のデータマイニング手法を流用した多次元構造分析
 - ドリルダウン、スライス&ダイス…

○ テキスト部分の内部分析との複合化

- 指定された単語の出現有無で頻度分析
- 係り受け、共起などの関係について、1文章中にどれくらいの%で同時出現するか、その関係をネットワーク図示
- 単語と単語、単語と属性との関係からクロス表 コレスポネンス分析
二次元チャート
- 文章間の距離を定義してクラスタ分析
 - 各文章につき、指定したN個の単語について、各文がそれを含んでいれば $d(i) = 1$ 、含んでなければ $d(i) = 0$ とするN次元ベクトルを定義し、文章同士の距離を計算し、クラスタ分析
- その他…

テキストマイニングのプロセス

ステップ4 プレゼンテーション

- 分析結果を視覚化、プレゼンする機能
 - 関係性の把握
 - ネットワーク図
 - グラフ機能
 - 1次元: 単純な頻度グラフ、ドリルダウン(属性別)の表示…
 - 2次元: 知覚マップ、散布図、…
 - 他ソフトとの連携
 - オフィス系ソフト
 - WEB形式
 - BIのためのコクピット設計

TMS活用事例

大学教育および研究ツールとしてのTMS

1. 受託リサーチ

- 商店街リサーチ (立教大学学部ゼミ)

2. 教育

- eマーケティング実習 (多摩大学大学院MBAコース)

3. 研究ツール

- 産業ネットワーク図

1. 受託リサーチへの適用

- 商店街のマーケティングリサーチ案件
- 自由意見欄で地域住民の意見を収集
- TMSで分析

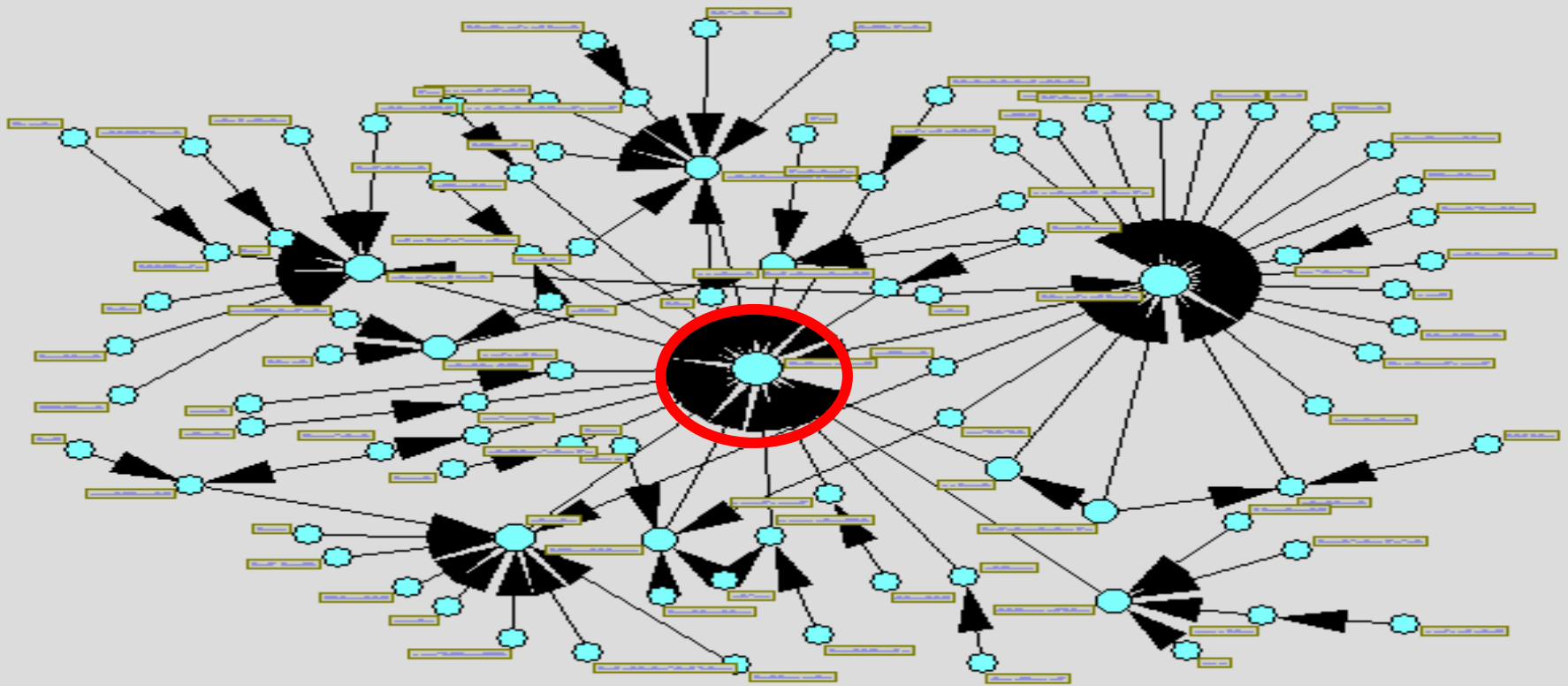
2. eマーケティング実習

- 多摩大学大学院MBAコース
- eマーケティングの実習
 - データマイニング : 統計処理
 - テキストマイニング : TMS
- 2005年後期より開始
 - 10月31日
 - 11月14日 2回で終了

3. 産業ネットワーク図

- ネットワーク図示にTMSを活用
 - IT産業の企業間元請け - 下請け構造分析
 - テキスト部分を『A社のB社』として記述
 - A社:発注先(下請け)企業
 - B社:発注元企業

某企業の下請け構造



おわり

○ 参考文献

- 拙稿(2005),「ソシオ・エコノミック統合型システム - オープンソース時代の情報システム設計に向けて」,OA学会第50回全国大会予稿集,pp.53-56.
- 拙稿(2005),「コンテキスト・マイニング - その対象とBI構築の方向性 - 」,OA学会学会誌, Vol.26-2(現在印刷中).
- 阿久津聡,石田茂(2002).「ブランド戦略シナリオ」,ダイヤモンド社.