

# 言語情報研究室

～大量のテキストデータから重要な情報を抽出～

酒井 浩之

## 背景

膨大な量のテキストデータ（**ビッグデータ**）の中から、目的に合う有用な情報を抽出するための技術（**自然言語処理**+**データマイニング** = **テキストマイニング**）を研究

### テキストマイニングの適用例：

- ✓ 評判情報自動分析（アンケート結果、Web掲示板、ブログを情報源として）
- ✓ 企業のコールセンターに寄せられる大量の質問事項を自動的に分析（新商品開発に使用）
- ✓ 企業の**決算短信**、**有価証券報告書**を自動的に分析（株式投資判断の支援に使用）

## 研究室の成果

### ■ 言語情報研究室におけるテキストマイニングの研究事例の紹介

- ✓ 企業の決算短信をマイニングした株式投資支援システム  
→ 決算短信から抽出した業績要因を検索対象とした決算短信検索システム
- ✓ 企業WEBページをマイニングした就職活動支援システム  
→ 企業WEBページから抽出したキーワードを使用した関連企業検索システム
- ✓ 大学WEBページをマイニングした志望大学選択支援システム  
→ 大学WEBページから研究室トップページを抽出した大学研究室検索システム

※ 言語情報研究室WEBサイトにて公開されており、利用可能

URL: <http://www.ci.seikei.ac.jp/sakai/>

## 応用例

例：「**太陽電池**」を業績要因とする企業の決算短信を検索

### 決算短信検索システム CEES

企業名を選択で、その企業Webサイトを閲覧可能

業績要因を表示

業績要因によく出現する企業キーワード

企業Webサイトで公開している決算短信を閲覧可能

