

# 「ツイちえき！」開発での取り組みについて

NTTコミュニケーションズ株式会社  
先端IPアーキテクチャセンタ  
大角知孝

2010/09/04

# 本日本話すること

- 「ツイちえき！」とは？
- 「ツイちえき！」でのワード抽出
- 「ツイちえき！」でのツイート取得
- 「ツイちえき！」開発の今後

# 「ツイちえき！」とは？

# 「ツイちえき！」とは

Twitterで、今、この瞬間に話題に注目されている話題をリアルタイムに抽出してお知らせするサイト。

最新の注目ワード

過去の話題を探す

ツイッターの直接検索

おすすめワード

関連ページリンク

関連フレーズ

ツイート一覧

くわしくはこちら  
<http://tweet.ocn.ne.jp/guide/>

# 「ツイちえき！」でのワード抽出

# 「ツイちえき！」でのワード抽出

- 単語ではなく、**意味のあるフレーズを単位**として抽出したい
- 自動抽出だけでなく、運営側が手動でおすすめワードも入れたい

注目ワードTOP30



自動抽出でのワード抽出

ツイちえき！おすすめワード



おすすめワードを手動登録して抽出

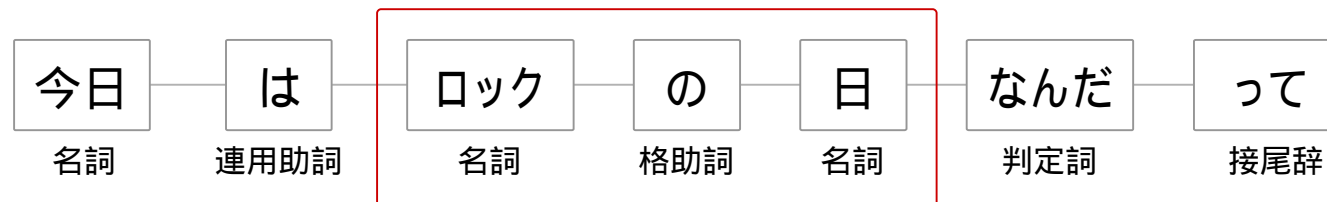
いかに綺麗にフレーズを認識して、  
抽出できるかがポイント！

# 自動抽出 - 「ツイちえき！」でのワード抽出

- 形態素解析での品詞列に注目して自動フレーズ抽出
- **意味のあるまとまり**として抽出することで話題をわかりやすく

## 【事例】自動抽出

「今日はロックの日なんだって！ <http://www.aaaaaa.com/bbbbb/>」



「ロック」ではなく「**ロックの日**」で抽出をする

NTTサイバースペース研究所が開発した日本語解析エンジンを利用



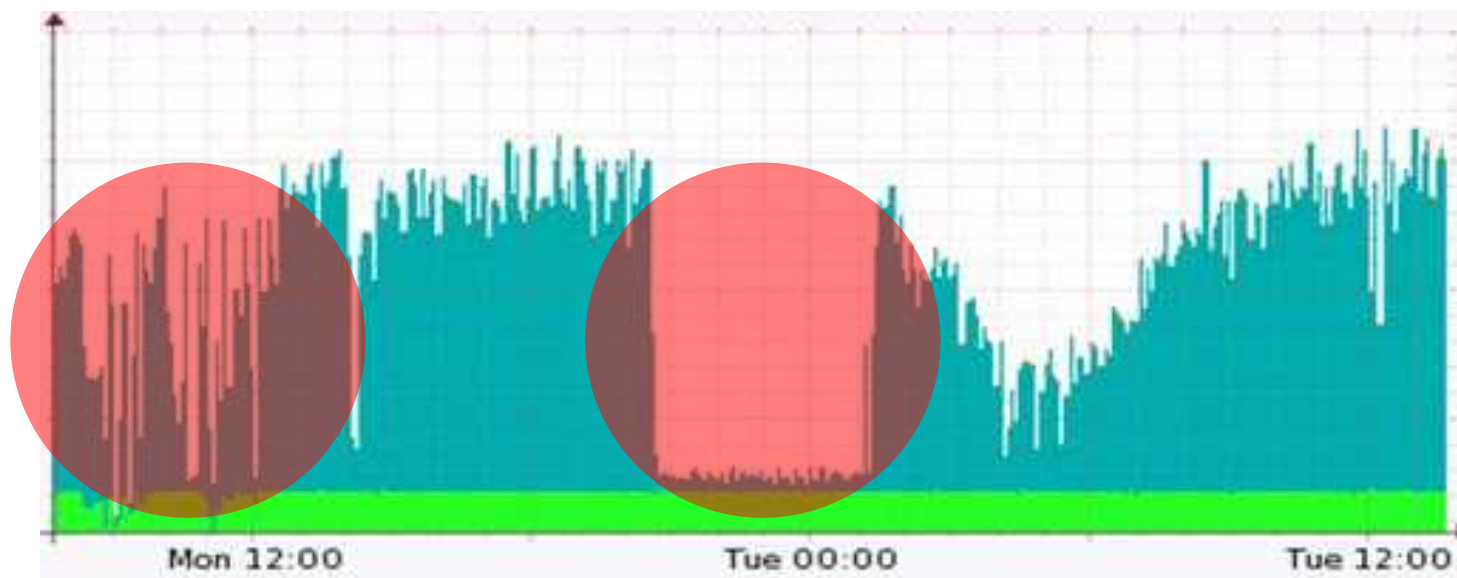


# 「ツイちえき！」でのツイート取得

# 「ツイちえき！」でのツイート取得

- 安定してツイートを取得することがサービス提供での鍵
- 注目されているワードを抽出するためにある程度**大量のツイートが必要**

【例】6/28(月)Search APIでのツイート取得  
オランダVSスロヴァキア戦の日



# 「ツイちえき！」でのツイート取得

- APIが不安定になることを想定した実装によるシステムトラブル回避を実現
- 並列して異なるAPIを利用することでツイート取得を安定的に

## 1. Streaming API (gardenhose)

- ツイート総数の約5%をほぼリアルタイムにストリーム配信
- 取得したツイートを順次解析していく

## 2. Search API (search.twitter.com)

- 検索結果として取得できるツイートを解析

# 「ツイちえき！」開発の今後

## 「ツイちえき！」開発のこれから



- スマートフォンへの対応
- より美しいワード抽出へ
- より旬な話題を取り上げられるように

これからも「ツイちえき！」にご期待ください

# 「ツイちえき！」公式アカウント



# @twichk



<http://tweet.ocn.ne.jp>

**ツイちえき!** 今この瞬間の話題を追跡!