

# 「日本語処理解析性能評価」分科会 活動報告及び活動予定について

平成29年4月20日

第13期主査 絹川 博之(東京電機大学)

1

## 「日本語処理解析性能評価」分科会

- 本分科会設立にいたる課題

デジタル・フォレンジック や eディスカバリ を用途とする  
検索・解析ツールに関し、海外で開発されたものが多く、  
日本国内での使用時に、日本語への対応が不明で、  
下記のような問題が発生！

- 検索モレがあり、重要証拠ファイルの不検出
- 文字化けの発生で、内容の確認不可
- どこまで日本語処理対応しているか不明確
- 日本語処理の精度の信頼度に課題
- 日本語処理対応の定義が不明確

2

# 「日本語処理解析性能評価」分科会

- 本分科会設立の目的

- デジタル・フォレンジック や eディスカバリ の対象となる日本語情報に対する処理解析性能を評価するための有効な指標を作成し、客観的な評価の実施を可能とする
- ツール提供企業の技術進歩を促し、デジタル・フォレンジック技術の日本国内でのさらなる発展に寄与する。

3

---

## 「日本語処理解析性能評価」分科会メンバー

(敬称略)

主査: 絹川博之(東京電機大学)

幹事: 野崎周作((株)FRONTEO)

幹事: 白井喜勝((株)FRONTEO)

### <メンバー>

舟橋 信((株)FRONTEO)

青嶋信仁((株)ディアイティ)

伊藤文二(日本ダイレックス(株))

春山 洋(AOSリーガルテック(株))

緒方 健(おがたコンサルティング)

栗原 葵((株)FRONTEO)

丸野修平((株)FRONTEO)

岡田 忠((独)情報処理推進機構)

藤本隆三(AOSリーガルテック(株))

青木和哉(FTI Consulting)

### <オブザーバー(個人の立場でご参加)>

野本靖之 石崎 俊 岡野 薫

4

# 2016年度の活動日程

- 2014年度「日本語処理解析性能評価」分科会：
  - 評価項目案の策定
- 2015年度「日本語処理解析性能評価」分科会：  
2015年4月～2016年3月 7回実施
  - 評価基準、評価データ、評価手順の作成
- 2016年度「日本語処理解析性能評価」分科会：  
2016年4月～2017年3月 6回実施
  - 評価委員会の設立と受検企業の募集
  - 製品性能評価の実施
  - 結果を IDF Webサイトで2017年3月より公表

5

# 2016年度の活動内容

- 日本語処理解析性能評価用データ：下記 5テキスト
  - IDF理事会承認と執筆者のご了承のもと、下記 3コラム使用
    - (1) コラム319号：町村泰貴理事  
「電子証拠の取扱いと訴訟法の違い」
    - (2) コラム345号：須川賢洋理事  
「著作権法の改正と電子出版の今後の展望」
    - (3) コラム360号：手塚 悟理事  
「IoT/M2Mの問題点」
  - 評価基準項目検証に必要な追加訂正を319号、345号に加えたもの
- 受検企業に対する評価実施説明会の実施  
株式会社FRONTEO、株式会社ディアイティ
- 立会人の下、製品評価の実施  
結果：IDF Webサイトで 2017年3月より公表

6

# 評価項目と性能評価

## (1) 文字コード: 9 種

UTF-8、UTF-7、UTF-16LE、UTF-16BE、Windows-31J、  
JIS、JIS-Allow 1 byte Kana、JIS-Allow 1 byte Kana-SO/SI、EUC-JP

## (2) アプリケーション: 5種

プレーンテキスト: 45種(5テキスト×9種コード)、  
Eメール: 6種、MS-Word 2013(17拡張子)、  
MS-Excel 2013(23拡張子)、MS-PowerPoint 2013(26拡張子)

## (3) 基本検索: 16項目

単一単語検索、ブーリアン検索、数字・漢数字同一視検索、etc.

## (4) 応用検索: 18項目

正規表現、近傍検索、片仮名・英数字の全角半角同一視検索、  
片仮名と漢字・平仮名の同形異字混在検索、etc.

## (5) 正答表の作成:

## (6) 評価指標:

再現率、適合率、正解率

7

---

## 2017年度の活動予定

### 「日本語処理解析性能評価」ワーキンググループ

#### • 体制

- 座長: 野崎周作様((株) FRONTEO)、  
幹事: 白井喜勝様((株) FRONTEO)、  
ワーキングメンバー、アドバイザーメンバー
- 評価委員会: 有識者メンバー

#### • 開催時期・内容

- 半期に1度製品評価を実施
- 評価基準、評価データの補正検討

8