

「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28年4月22日

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)

幹事: 野崎 周作(株UBIC)

幹事: 白井 喜勝(株UBIC)

<メンバー>

舟橋 信(株UBIC)

栗原 葵(株UBIC)

青木 和哉(FTI Consulting)

青嶋 信仁(株ディアイティ)

伊藤 文二(日本ダイレックス株)

岡田 忠(独立行政法人情報処理推進機構)

春山 洋(AOSリーガルテック株)

藤本 隆三(AOSリーガルテック株)

緒方 健(おがたコンサルティング)

<オブザーバー(個人の立場でご参加)>

石崎 俊 岡野 薫

4

2015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
 - 評価項目の具体内容の決定
 - 評価用データの作成
 - 評価を実行するための
日本語処理解析性能評価委員会の
募集要項の作成

5

2015年度の活動内容

- 日本語処理解析性能に関する評価用データの選定
 - IDF理事会承認のもと、コラム記事より選定
 - 執筆者のご了承の頂戴
 - 下記コラムを使用させて頂く
 - (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」
- 評価項目の決定と評価用データの作成
- 評価委員会募集要項の作成

6

評価項目の決定と評価用データの作成

- (1) 文字コード: 9 種 (<= 8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: 6+4 種 (<= 7+5種)
Eメール: Outlook系4種、Becky!、Thunderbird
プレーンテキスト、MS-Word 2013(17拡張子)、
MS-Excel 2013(24拡張子)、MS-PowerPoint 2013(28拡張子)
- (3) 基本検索: 2種
単一単語検索及び2単語のブーリアン検索(AND、OR、NOT)について
可否確認要として詳細化
- (4) アドバンス検索: 6種 (<= 5種)
正規表現検索、前後に特定文字が来ない検索、グルーピング、近傍検索、
全角・半角の片方入力で自動的に両方検索、表記のゆらぎ
- (5) 評価用データの作成:
3種のコラム原文に対して、(1)-(4)の評価に使用するキーワードを検討し、
評価用データを作成。(1)-(4)に適宜フィードバックし、項目を再設定。

7

「日本語処理解析性能評価」分科会 2016年度の活動予定

- 日本語処理解析性能評価委員会の設置
 - 評価委員会の募集要項に基づき委員の募集
 - 評価委員会の立上げ
- 分科会: 5月から、隔月開催
- メンバー
 - 第12期(2015年度)に同じ。
必要により逐次メンバーを追加。
 - アドバイザーとしてオブザーバー参加頂く方々(個人)
石崎 俊、岡野 薫、野本 靖之
有志による参加者を逐次追加。

8