

## 第3回「日本語処理解析性能評価」実施結果報告

### 1 評価結果の客観的評価指標

#### (1)「日本語処理解析性能評価」実施の目的(趣旨)と評価の実施について

近年、デジタル・フォレンジックやeディスカバリ用途で、多様な検索機能や解析機能を持つ多くのソフトウェアが開発され、利用されています。しかしながら、海外で開発されたものも多く、日本国内で使用する場合に、どこまで日本語に対応しているかが不明で、ユーザーが使用してみるまでわからないのが現状です。また、実際に性能を評価しようとしても、客観的かつ有効な評価基準や指標も存在しないため、比較自体が困難な状況です。IDF「日本語処理解析性能評価」分科会ではこの状況を改善するために、日本語処理解析性能を評価するための基準となる項目とそれに伴う検索クエリ、さらに実際の評価に使うための評価用データの作成と各種ツールの日本語処理解析性能の評価基準項目を準備しました。また、本評価を実現するために日本語処理解析性能評価委員会が設立され、2017年1月に第1回評価が実施されました。

今期第14期第2回目(通算第3回目)の評価は、今期計画に従いオープンソースソフトウェアの「Autopsy」の評価を実施致しましたので結果を報告致します。

尚、「Autopsy」は、Basis Technology社の製品ですが、日本では同社日本法人を含め販売やサポートを行っている企業がないため、「日本語処理解析性能評価」WGの野崎座長の所属する(株)FRONTEOにてダウンロードした「Autopsy」を使用して評価の受検準備と実施を担当致しました。

#### (2) 評価基準と評価用データ作成について

##### 評価基準について

単一単語検索や簡単なブーリアン検索を中心とした「基本検索」だけでなく、全角半角の同一視検索、正規表現や近傍検索といった比較的高度な検索機能を含む各種の「応用検索」も評価基準に含めました。尚、単一単語検索の項目にも「数字・漢数字同一視検索」といったものも含まれているので、「基本検索」といっても難易度は高いと想定されます。さらにそれぞれの評価用データとしては日本語特有の文字コードが多数と日本特有のEメールソフト等も対象に含みますので、総じて難易度の低いものから非常に高いと思われるものまで幅広く評価が可能なものとなりました。

##### 評価用データ作成手順について

評価用データは、IDFコラムより319号、345号、360号のそれぞれのテキスト情報と、各種評価基準項目を検証する為に必要な追加修正を319号、345号に加えたものの合計5種類のテキスト情報をベースに、9種類の文字コードのテキストファイル、テキストファイルを添付した6種類のEメールアプリケーションデータ、Microsoft Office (Word/Excel/PowerPoint) を用いた各種保存形式の違いによる多数のファイル形式の評価用データを作成しました。

### 2 評価結果の客観的評価指標

各種評価基準項目ごとに実施しました評価用データに対する検索結果は、評価用データに「ヒットしたか」「ヒットしなかったか」、それはそれぞれ評価用データにおいて「正解としてヒットすべきものか」「正解としてヒットすべきでないか」という2つの軸が存在し、【表1】で表される4象限で表記できます。

【表1】評価結果の4象限

		評価用データにおける正解	
		ヒットすべきもの	ヒットすべきでないもの
検索結果	ヒットした	True Positive (TP)	False Positive (FP)
	ヒットしなかった	False Negative (FN)	True Negative (TN)

統計分析で良く使われる指標として、下記3つの指標がありますので、今回の評価結果を示す客観的数値として公表致します。

(1) Recall : 再現率

正解としてヒットすべきもののうち、検索で正しくヒットしたものの割合

$$\text{Recall : 再現率} = \frac{\text{TP}}{\text{TP+FN}}$$

(2) Precision : 適合率

検索でヒットしたもののうち、正解としてヒットすべきものの割合

$$\text{Precision : 適合率} = \frac{\text{TP}}{\text{TP+FP}}$$

(3) Accuracy : 正解(答)率

検索でヒットしたものとヒットしなかったもの(全体)がそれぞれ正解である割合

$$\text{Accuracy : 正解(答)率} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}}$$

今回の「Autopsy」の評価結果を【表2】に記載します。

【表2】「Autopsy」評価結果

テキスト 基本検索		評価用データにおける正解		合計	適合率
		ヒットすべき	ヒットすべきでない		
検索結果	ヒットした	90	1	91	99%
	ヒットしなかった	186	293		
	合計	276		正解率	67%
	再現率	33%			

テキスト 応用検索		評価用データにおける正解		合計	適合率
		ヒットすべき	ヒットすべきでない		
検索結果	ヒットした	90	0	90	100%
	ヒットしなかった	279	441		
	合計	369		正解率	66%
	再現率	24%			

Eメール 基本検索		評価用データにおける正解			
		ヒットすべき	ヒットすべきでない	合計	適合率
検索結果	ヒットした	10	0	10	100%
	ヒットしなかった	2	12		
	合計	12		正解率	92%
	再現率	83%			

アプリケーション 基本検索		評価用データにおける正解			
		ヒットすべき	ヒットすべきでない	合計	適合率
検索結果	ヒットした	38	0	38	100%
	ヒットしなかった	26	2		
	合計	64		正解率	61%
	再現率	59%			

各テーブル上でカウントされている数は評価用データにおける検索ヒット数ではなく、ヒットファイル数になります。

### 受検製品「Autopsy」のコメント、問い合わせ先等

	受検製品：「Autopsy」	Ver.4.6.0
製品特性	Basis Technology 社WEBサイト ( <a href="https://www.basistech.jp/">https://www.basistech.jp/</a> ) から引用 Autopsy®は、オープンソースのデジタルフォレンジックプラットフォームです。商用フォレンジックツールに必要なとされる中核機能を搭載した Autopsy は、どんなニーズにも応えられる高速で完全かつ効率的なハードディスク診断ソリューションです。	
評価実施目的	代表的OSS製品の日本語処理解析の客観的な性能を知る。	
評価結果へのコメント	国産製品や、海外製品でも日本向けにカスタマイズされ、販売・サポートされている商用フォレンジックツールと比べると低い正解率となったが、今回の評価結果を元に性能を正しく把握すれば、実際の日本語データに対するフォレンジック調査においても十分活用できるツールである。	
問い合わせ先	I D F 事務局 〒141-0022 東京都品川区東五反田 1 - 4 - 1 八二一五反田第2ビル3F TEL : 03 - 5420 - 1805 FAX : 03 - 5420 - 3634	
Autopsy について	WEBサイト ( <a href="https://www.autopsy.com/">https://www.autopsy.com/</a> ) ご参照	