

第1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、第1回評価が実施されました。

(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

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

今回の株式会社 FRONTEO (Lit i View E-DISCOVERY) 及び株式会社 ディアイティ (X-Ways Forensics) の評価結果を【表2】【表3】に記載します。

【表2】株式会社 FRONTEO (Lit i View E-DISCOVERY) 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	222	0	222	100%
	ヒットしなかった	54	294		
合計		276		正解率	91%
再現率		80%			

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	297	0	297	100%
	ヒットしなかった	54	459		
合計		351		正解率	93%
再現率		85%			

Eメール 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	12	0	12	100%
	ヒットしなかった	0	12		
合計		12		正解率	100%
再現率		100%			

アプリケーション 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	59	0	59	100%
	ヒットしなかった	7	0		
合計		66		正解率	89%
再現率		89%			

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。

【表3】株式会社ディアイティ (X-Ways Forensics) 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	191	15	206	93%
	ヒットしなかった	85	279		
合計		276		正解率	82%
再現率		69%			

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	181	19	200	91%
	ヒットしなかった	170	440		
合計		351		正解率	77%
再現率		52%			

Eメール 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	12	0	12	100%
	ヒットしなかった	0	12		
合計		12		正解率	100%
再現率		100%			

アプリケーション 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	53	0	53	100%
	ヒットしなかった	13	0		
合計		66		正解率	80%
再現率		80%			

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。

受検社（株）FRONTEO及び（株）ディアイティのコメント、問い合わせ先等

株式会社FRONTEO	受検製品：Lit i View E-DISCOVERY	Ver.7.12.203548
製品特性	<p>本製品は、eディスカバリをはじめとする大容量データのフォレンジック調査を効率的に行うことを目的に開発された製品です。日本語をはじめとするアジア言語の解析に優れ、ブーリアン検索から近傍検索、正規表現検索などアジア言語に対しても正確に検索することが出来ます。また、人工知能KIBITを搭載し、少量の証拠データをKIBITに学習させることによりKIBITが同様の証拠データを見つけ出すサポートを行います。</p>	
受検目的	本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	評価対象となっている全ての文字コードに対応でき、個々の評価項目に対して満足できる結果が得られた。	
問い合わせ先	株式会社FRONTEO リスクコンサルティング部 TEL：03-5463-7577	
自社HP公開	FRONTEO WEBサイト (http://www.fronteo.com/corporate/news/2017/20170315130000.html) にて、本評価結果に関する自社コメントの公開しております。また、詳細結果等の確認をご希望の場合は、上記問い合わせ先にご連絡下さい。	

株式会社ディアイティ	受検製品：X-Ways Forensics	Ver.19.0 SR-12
製品特性	<p>本製品は、フォレンジック調査を目的としたツールで、記憶媒体の保全から、対象情報の解析や復元まで行えます。豊富な機能、メニューの使い易さと動作の軽快さと共に、日本語対応のマニュアルとヘルプがあり、捜査機関やフォレンジックサービス業者で広く使われています。</p> <p>バイナリエディタとして世界的に定評のある「WinHEX」をベースとして開発されており、ファイルの復元機能、レジストリレポート機能、タイムライン生成機能等々フォレンジック調査に必要な多くの機能を搭載しています。</p> <p>それらを利用するためのユーザーフレンドリーなGUIが使い易く、初心者からエキスパートまで非常に好評を得ています。またデータの保全、抹消も可能であるため一通りのフォレンジック解析を1本のソフトで完了させることができます。さらに、データベースの準備やソフトウェア自体のインストールを必要とせず、USBメモリからの起動も可能であるため、ライブフォレンジックや調査対象のマシンを絞り切れないときの表層解析等にも非常に有効です。</p>	
受検目的	日本語処理解析を求めている製品ではないが、その目的で使った場合に、どのくらいの性能があるかを知るため。	
評価結果へのコメント	大体想定どおりだった。フォレンジック調査を目的とする種類のツールにおいての日本語処理解析性能は、大体同様の結果となります。	
問い合わせ先	株式会社ディアイティ セキュリティサービス事業部 TEL：03-5634-7654	
自社HP公開	自社サイトでのコメントは掲載致しません。上記へお問い合わせ下さい。	