

第15期第3回(通算6回目)「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15期第3回目(通算第6回目)の評価は、2019年3月に(株)くまなんピーシーネットの「Intella」(2017年12月に評価受検した製品のバージョンアップ版: IDFのHPに評価結果を公開済み)の評価を実施致しましたので結果を報告致します。

尚、これまでの6回の評価実施で得られた知見に基づき日本語処理解析性能評価委員会では、第16期に「評価基準」及び「評価用データ」についての補備・修正検討を行い、より適正な評価を実施できるように致します。第16期の評価は、この補備・修正された「評価基準」及び「評価用データ」で実施致します。

(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

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

今回の株式会社くまなんピーシーネット (Intella Professional) 評価結果を【表2】に記載します。

【表2】株式会社くまなんピーシーネット (Intella Professional) 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	220	0	220	100%
	ヒットしなかった	56	294		
合計		276		正解率	90%
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	72	441		
合計		369		正解率	91%
再現率		80%			

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	5	2		
合計		64		正解率	92%
再現率		92%			

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）くまなんピーシーネットのコメント、問い合わせ先等

株式会社 くまなんピーシーネット	受検製品：Intella Professional	Ver.2.2.1
製品特性	<p>Intella は、Office や Acrobat に代表されるドキュメントファイルやメール、グループウェア等のデータから、ファイルに含まれる目的の文字列を検索し、様々な視覚的表現で検索結果を表示できる優れたフォレンジックサーチソフトウェアです。</p> <p>これら検出能力を支える Intella の強力な解析エンジンは、インストールされたパソコンの性能にほとんど負担を与えず、低スペックのパソコンから高性能なパソコンまで変わらぬ検索結果を提供します。ライセンスについても、あらゆるユーザーが導入しやすいラインナップで用意されており、個人や小規模オフィス、フォレンジック専門家から大勢で行う解析まで用途や規模で選択できます。また、各バージョンからのアップグレードもいつでも簡単に行えるため、解析容量の変化に柔軟に対応できます。</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	<p>今回で2回目の検定となります。前回の検定結果が公開されて以降、様々な業界のお客様から多数のお問い合わせを頂き、IDFによる「日本語処理解析性能評価」は、日本の水準となるべき検定基準と受け止めています。</p> <p>今回は、前回の検定で課題となったアプリケーションのヒット漏れに特化した改善を行い再受検しました。大きなスコアアップまでには至りませんでした。前回の課題をクリアできたことで製品をご利用中のユーザーや現在導入をご検討中のユーザーに対し、製品の信頼面を証明する成果となったことに間違いありません。今後さらなる改善に挑んでいく所存です。</p> <p>最後に、2回の検定を経て受検者の立場として要望を申し上げます、「製品が持つ固有の日本語解析能力（製品独自の機能）」等の評価基準を設けて欲しいことや、<u>検定に使用するパソコンは IDF でご用意頂き「パソコンの性能や OS のバージョンなどを統一した評価」にして、本結果をご覧になった方々が判断しやすい検定環境要素も取り入れて頂きたいと望んでいます。</u></p>	

問い合わせ先	株式会社くまなんピーシーネット 〒860-0834 熊本県熊本市南区江越2 - 1 - 8 TEL : 096 - 373 - 2213 FAX : 096 - 373 - 2214
自社HP公開	WEBサイト (https://www.kumanan-pcnet.co.jp) にて、本評価結果に関する自社コメントの公開を予定しております。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。

ご要望（コメントでの下線部分）へのご回答	<p>本報告書の先頭に記してございます経緯の通り、本評価を具体化し実施するために日本語処理解析性能評価委員会が編成され約2年に渡り準備と検討を進めた際にご要望頂いた点も検討したのですが、ツールによってPC等のオンプレミスで稼働するソフトもあればクラウド環境で使用するソフトもあるので受検企業が最適と考えるスペックの使用環境を自ら評価に提供するという形で評価を実施するという結論となりました。また、評価に準備する使用環境のスペックによって処理時間等のスピードに差が出てくることも想定されたため、受検企業が評価受検に当たり、高いスペックのPCを準備できないから受検を控えようということにならないよう評価項目から処理時間の項目は除外するという結論となり、その様に実施して来ております。</p> <p>ご理解をお願い致します。</p>
----------------------	---

以上