

第19期第1回(通算第8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。

日本語処理解析性能評価委員会では、第17期、第18期の間にWGを複数回実施し、これまで4期にわたって実施してきた評価を振り返り、修正が必要な個所及び要望の多かった評価項目の追加に関し検討を行い、「評価基準」と「評価用データ」の改訂を実施しました。

改訂された「評価基準」と「評価用データ」にて実施する今期第19期第1回目(通算第8回目)の評価は、2023年3月に(株式会社FRONTEO)の「Lit i View E-DISCOVERY」(※2017年3月に評価受検したLit i View E-DISCOVERYのバージョンアップ版)の評価を実施致しましたので結果を報告致します。

(2)評価基準と評価用データ作成について

①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。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。

②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7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

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

(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

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

(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

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

今回の株式会社 FRONTEO (Lit i View E-DISCOVERY)評価結果を【表2】に記載します。

【表2】株式会社 FRONTEO (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	204	0	204	100%
	ヒットしなかった	0	186		
合計		204		正解率	100%
再現率		100%			

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	279	0	279	100%
	ヒットしなかった	9	432		
合計		288		正解率	99%
再現率		97%			

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	21	0	21	100%
	ヒットしなかった	0	21		
合計		21		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	66	0	66	100%
	ヒットしなかった	6	0		
合計		72		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 株式会社 FRONTEO 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.7.23 SP 1.234886
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 が類似の証拠データを見つけ出すサポートを行います。</p>	
日本語マニュアル	日本語マニュアルがございます。	
日本語インターフェース	日本語のユーザインターフェースでの利用が可能です。	
表記揺らぎ	<p>例示にあるような表記揺らぎの検索機能は搭載しておりません。 例1)デジタル・デジタル, バイオリン・ヴァイオリン等の表記の揺らぎを 同一視して検索 例2)カタカナ「カ」、漢字「力」など類似の表記を同一視して検索 例3)漢数字「二十一」と数字「21」を同一視して検索</p>	
人工知能	<p>弊社独自で開発を行いました人工知能 KIBIT を搭載しており、TAR(Technology Assisted Review)にも対応しております。また、弊社の KIBIT Automator と連携することにより、レビュー対象ドキュメントのスコアリングによる優先順位付けだけでなく、対象ドキュメントのカットオフやレビュー品質の QC 機能もご活用いただけます。</p>	
受検目的	本製品の日本語処理解析の客観的な性能を確認するために受検しました。	
評価結果へのコメント	<p>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ました。引き続き性能向上に向けて開発を進めていきます。</p>	
問い合わせ先	<p>株式会社 FRONTEO リーガルテック AI 事業本部 TEL:03-5463-7577 URL:https://legal.fronteo.com/contact/</p>	
自社HP公開	<p>自社 HP による公開は予定しておりません。 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。</p>	