

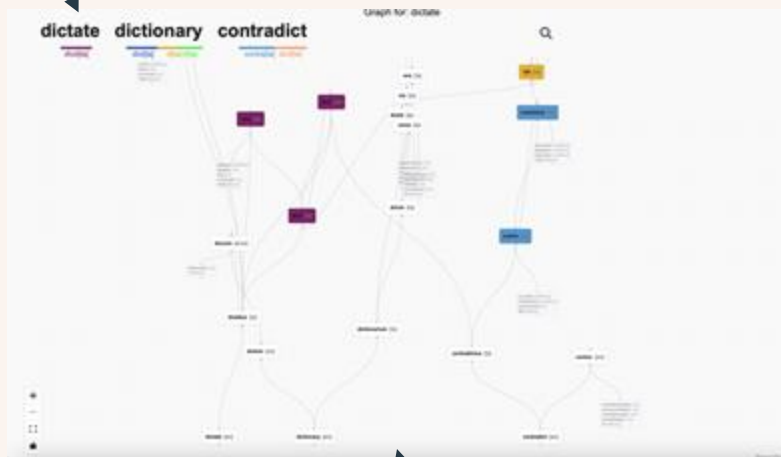
2024年度未踏IT人材発掘・育成事業

機械学習を用いた語源的英単語分割手法の開発 — 同語源単語を比較する学習アプリケーション —

中澤正樹（東京大学）

語源情報を生かした単語学習

スペルと語源を対応づけ



複数の共通語源単語を並べて表示

中級以上の単語や専門単語にも対応

語源的新語生成

Etymological word coinage (put two or three English words)

Submit

Results

Predictions

etymoronic etymoric chemomerical merological anger etemory morolitic
merochrome myrthicist ethemorometic etemor mormer mortication etemore
etymorome etemone

References

-ic[ia] mor[ang] *mur[gem-pro] *mora-[gem-pro] etymon[en] -tic[en] etymo[el]
*mer-[ine-pro] -ic[en] -icus[ia] etymos[el] *mor[gmw-pro] etymon[el] mor[enm]
-more[enm] -tic[de]

入力単語の先祖を探索し派生語を予測することで、
語源的に自然な架空の単語を作成

1.冗長性ある高品質な 語源ネットワーク

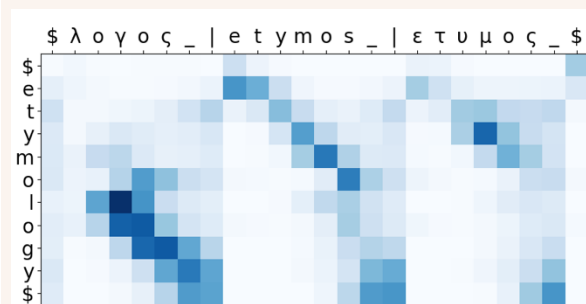
重複率

32.8 %

= 76.7万件 / 233万件

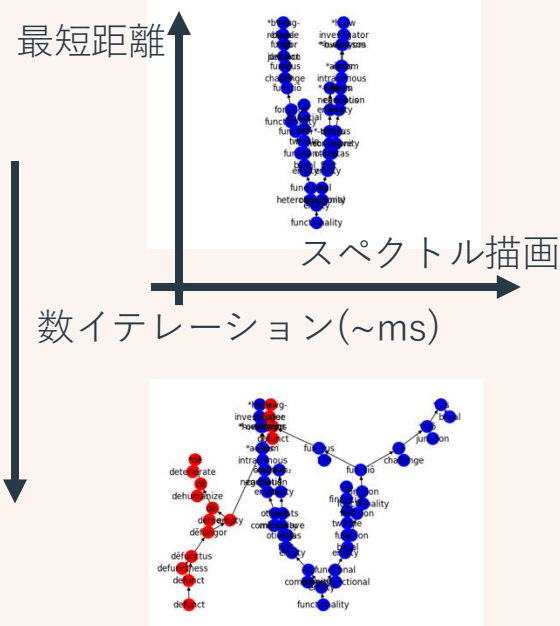
複数のデータセット・解析手法に
基づくことで、誤りの少ない
語源ネットワークを構築

2.スペルと語源を 紐づける語源的単語分割



Cross-Attentionから機械的に
派生語と語源語の間の
アラインメントを作成

3. 並べても崩れない 階層的グラフ描画



一次元方向でスペクトル描画を
採用することで逐次更新可能な
階層的描画