Empowered by Innovation NEC

【別紙】

利用者に応じた入力候補を表示する文字入力支援システムを開発

～キー操作でも音声でも自然な文字入力を実現～

2013年12月18日

NECソフト株式会社

従来の入力支援とその課題

- 利用者が入力したい情報を推定することで快適な入力環境を実現する**入力支援技術**が普及しつつある
  - キーワードの自動補完
  - スマートデバイスにおける音声入力
- しかし、利用者が日ごろ使う固有名詞や専門用語をシステムが知らないため、利用者の意図通りの入力ができない
  - 社内用語や商品名が入力できないと、ビジネス用途に使えない

© NEC Soft, Ltd. 2013

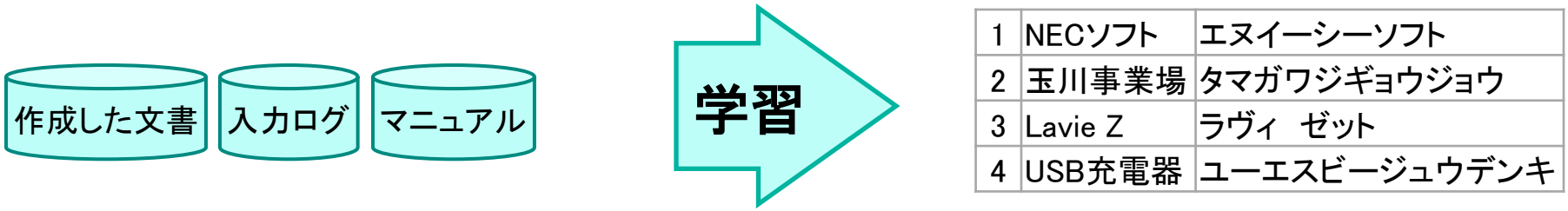
文字入力支援システム

1年以内に製品化予定

利用者が使う言葉を学習することで、意図通りの入力を実現

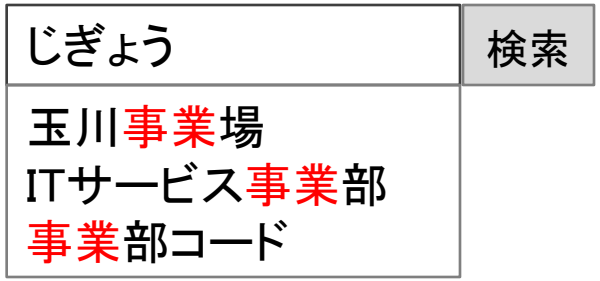
- 文書から利用者が使う言葉を学習し、入力支援の辞書を**自動生成**

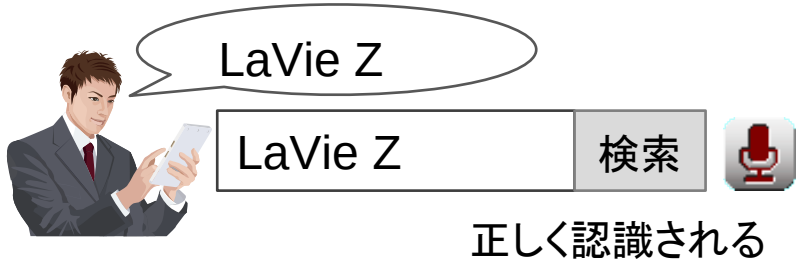
① 文書をクローリング　② 含まれる言葉を学習　③ 読みがなを推定



1	NECソフト	エヌイーシーソフト
2	玉川事業場	タマガワジギョウジョウ
3	Lavie Z	ラヴィ　ゼット
4	USB充電器	ユーエスビージュウデンキ

- 生成された辞書を使って、自動補完や音声入力を実現





正しく認識される

辞書を圧縮したまま高速に検索する技術により、省メモリ環境での自動補完を実用化

読みがな推定技術により固有名詞や特殊用語の音声入力が可能

© NEC Soft, Ltd. 2013

特長1　様々な文書形式から入力支援辞書の自動生成が可能

利用者が日ごろ使用している文書を分析し言葉を学習

- 過去に作成した文書や議事録、マニュアル、入力ログなどを分析
  - Microsoft® Officeドキュメント やPDFファイルなど、様々なフォーマットに対応
- 機械学習による分割で複雑な複合語も正確に抽出
- 使用状況に基づき重要度を測定し、入力支援時のランキングに利用

利用者がよく使う語をランキング

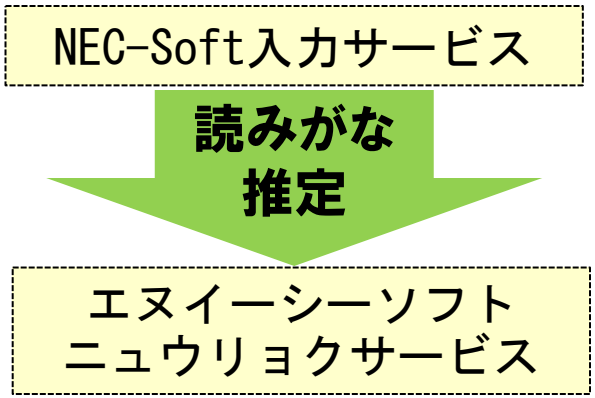
1	NECソフト	エヌイーシーソフト
2	玉川事業場	タマガワジギョウジョウ
3	Lavie Z	ラヴィ　ゼット
4	USB充電器	ユーエスビージュウデンキ
…	……	……
…	……	……

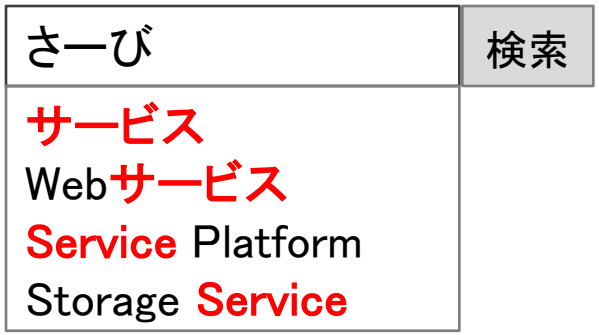
© NEC Soft, Ltd. 2013

特長2　読みがな推定技術により、文字形式に依存しない推薦候補の推定が可能

一般的な辞書にない言葉でも、アルファベット、漢字、カタカナの複雑な組合せでも、読みがなを推定できる

- かな漢字変換前に、先回りして変換結果を推薦できる
- 音声入力するための辞書として利用可能

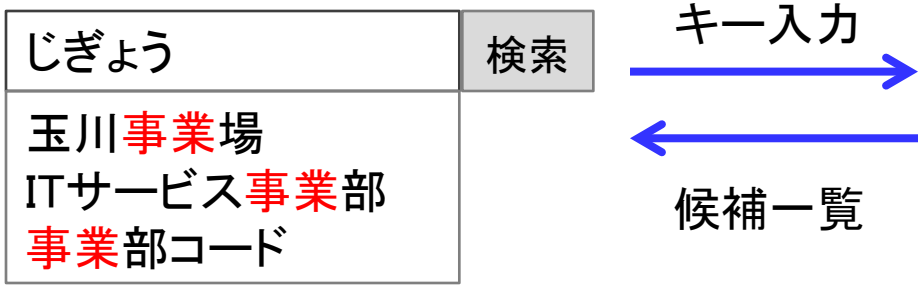




© NEC Soft, Ltd. 2013

特長3　省メモリでの高度な推薦を実現する圧縮検索技術

- 辞書を(最高で)元のサイズの10%に圧縮しつつ、利用者が次の文字を入力するよりも速く検索できる
- ランキングに基づき、利用者が入力しやすい語を優先して推薦
- 語の区切り位置を推定することで、正確なマッチングが可能



辞書

1	NECソフト	エヌイーシーソフト
2	玉川事業場	タマガワジギョウジョウ
3	Lavie Z	ラヴィ　ゼット
4	USB充電器	ユーエスビージュウデンキ
…	……	……
…	……	……
…	……	……

	レコード数	圧縮前	従来技術	NECソフト技術
URLリスト	32万	24MB (100%)	7.6MB (30%)	2.5MB (10%)
Wikipediaページタイトル	950万	195MB (100%)	85MB (44%)	60MB (31%)

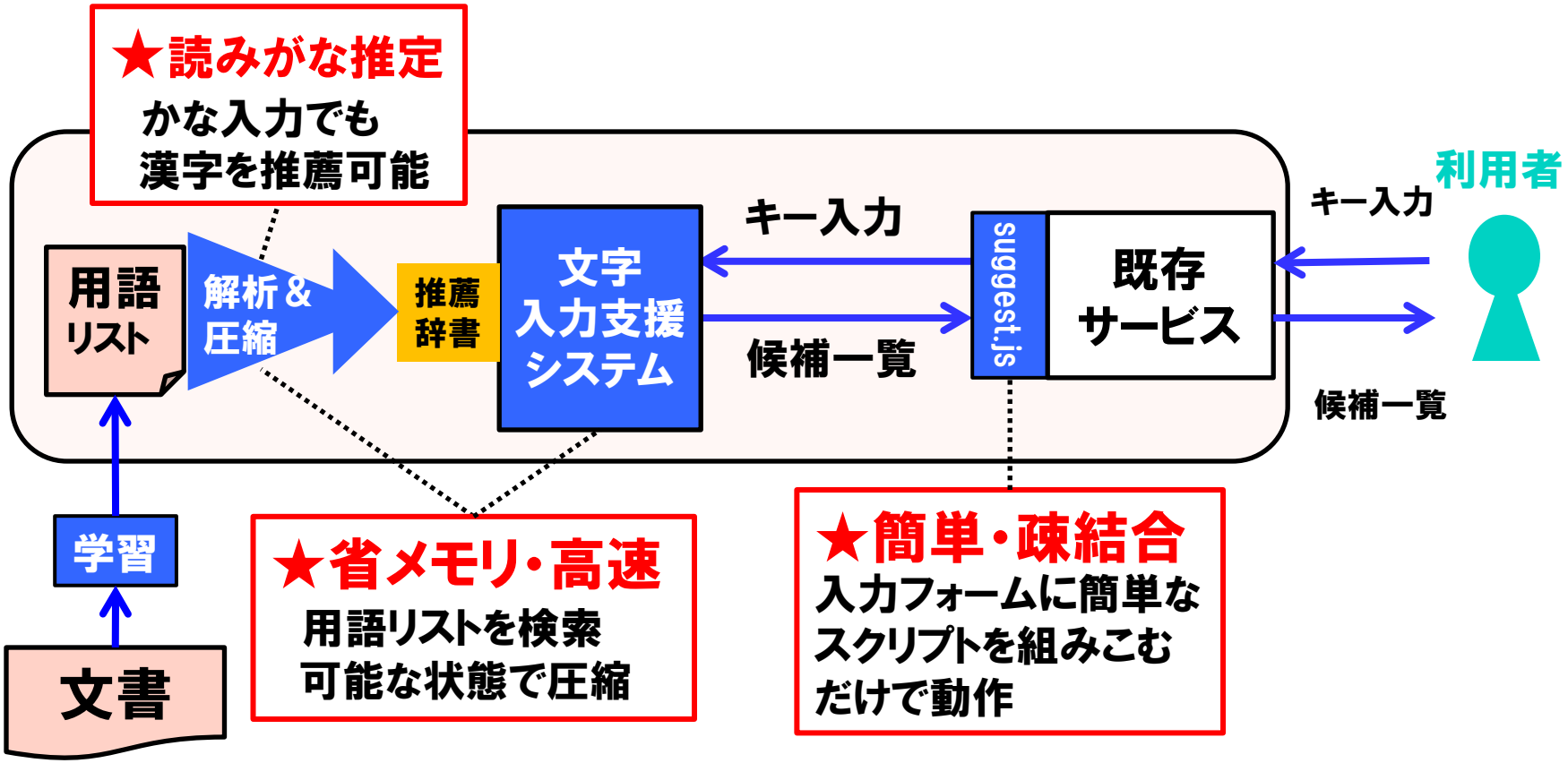
※カッコの中は、圧縮前とのサイズ比率
※従来技術：圧縮全文索引「FM-Index」
※研究成果の一部を国際会議「ALENEX2014」で発表予定

© NEC Soft, Ltd. 2013

特長4　既存システムへの導入が可能

既存サービスにJavaScriptを組み込むだけで簡単に導入可能

- 既存の企業向けの検索サービス、基幹系サービスなどに導入し動作を確認。実証評価中。

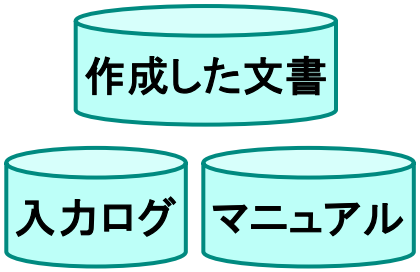


© NEC Soft, Ltd. 2013

動作イメージ（自動補完）

利用者が使う言葉の辞書を短時間で自動構築

① 文書を収集　② 言葉を学習　③ 読みがな推定



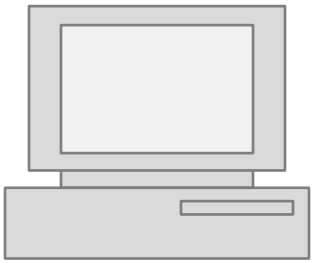
学習

1	NECソフト	エヌイーシーソフト
2	玉川事業場	タマガワジギョウジョウ
3	Lavie Z	ラヴィ　ゼット
4	USB充電器	ユーエスビージュウデンキ

じぎょう	検索
玉川事業場 ITサービス事業部 事業部コード	

キー入力

候補一覧



辞書を超圧縮し高速検索

文字を打つと
利用者が日ごろ使う言葉を推薦

© NEC Soft, Ltd. 2013

画面イメージ（自動補完）

© NEC Soft, Ltd. 2013

想定している利用シーン

これまで困難だった、利用者が日ごろ使用している言葉での入力が可能になります

- 販売員が自社の商品名を迅速に入力する。
- 会社員が社内用語を使用して容易にスケジュールを入力する。
- 医療現場において、手元が使えない環境で音声入力によって記録を取る。
- 小学生が入力する際、学校で習った語彙だけを表示させる。

© NEC Soft, Ltd. 2013

Empowered by Innovation