

Web 検索による語彙学習支援のアーキテクチャに関する考察

M2023SE008 大西翼

指導教員：野呂昌満

1 はじめに

Web 検索による学習促進のために加藤ら [1] は、空欄補充問題の自動生成手法を提案した。これは学習者に追加タスクを与え、自然な学習とは言い難い。我々は自然な学習とは学習者が問題を解く等の追加タスクなしで、無意識に学習できることを指している。

既存の検索システムにおける問題点は、自然な学習の補助を目的とする機能が搭載されていない点である。

本研究の目的は、自然な学習の支援である。すなわち、Web 検索による語彙学習において、習熟度向上を支援するアーキテクチャを提案する。検索時の使用者の操作において、手軽さを残した上で学習促進を目指す。自然な学習を促す方法として、ハイパーテキストから着想を得た。

研究課題として、以下の3点を挙げる。

課題 1. 学習促進方法の提案

課題 2. アーキテクチャの定義

課題 3. 妥当性の考察

本研究では、既存 Web ページに学習促進処理をしたコピーページを作成することを提案する。学習促進処理は、使用者に自然な学習を促す目的で、以下の処理をおこなう。

- 使用者習熟度を用いた検索対象語の特定
- 検索対象語のハイパーテキスト化

これらにより、使用者が疑問に思う語の追加検索を促す。

最後に、提案したアーキテクチャの設計について有効性、妥当性の観点から定性的に議論する。

2 先行研究・関連技術

2.1 先行研究

加藤ら [1] は、学習者による課題展開を促進するための空欄補充問題生成手法を提案した。個々の学習者による調べ学習の内容に応じて、文脈を考慮した問題を生成した。学習者の主体的な課題展開に対して、提案手法の有効性を確かめた。空欄補充問題を解くことで、調べ学習した内容の見直しが促され、新たな課題展開が促進される可能性を示している。問題と答えを与えることで学習促進を目指しているが、これは使用者の主体的で自然な学習といえるか懐疑的である。

2.2 ChatGPT

ChatGPT[2] は、OpenAI が開発した大規模な言語モデルで、膨大なテキストデータを使って学習している。質問応答、文章生成、会話等を行うことができる。文脈を理解し、次に来る言葉を予測する形で動作することで、人間のような自然な文章を生成する。学習は完了しており、その即時性から多くのアプリケーションで活用されている。

3 課題解決のアプローチ

本研究では、学習者が Web 検索を通じて自然と反復学習を行い、習熟度を向上させる仕組みの構築を目指す。

3.1 課題 1 へのアプローチ

学習促進処理の目的は、自然な学習の促進であり、その利点を以下で説明する。検索対象語とは、Web ページ内における重要語のことである。重要語とは、学習者が検索した際に閲覧する Web ページのテキストから検索した語に関連があり、学習の前提として、知るべき語句を対象としたものである。

使用者習熟度を考慮してこれを特定し、ハイパーテキスト化することには以下の利点がある。

- 検索対象語の可視化
- 追加検索の簡易化

検索対象語の可視化は、検索と同時に使用者への自然な学習を促すことが期待できる。Web ページ上で検索対象語に色付けにより、特定の語が強調表示され、使用者が学習時に目を向けるべき推奨箇所をシステムが注意喚起できる。特定箇所を使用者の習熟度によって変化させることで、その使用性、有効性の向上を目指す。また、使用者が疑問に思った語句を辿れる道すじができた点において、追加検索が簡易化されたといえる。加えて、リンク先への遷移によって、検索対象語について複数回の学習を行えるので、使用者に自然な反復学習を促すことが出来る。

例として、図 1 に、作成を想定する Web ページを示す。これは、Wikipedia の一部分 [3] で、提案するシステムが想定する出力である。

HTML^[注釈 1]または**HyperText Markup Language**（ハイパーテキスト マークアップランゲージ）^[注釈 2]は、ハイパーテキストを記述するためのマークアップ言語の1つで、プログラミング言語ではない。主にWorld Wide Web (WWW) において、ウェブページを表現するために用いられる。ハイパーリンクや画像等のマルチメディアを埋め込むハイパーテキストとしての機能、見出しや段落といったドキュメントの抽象構造、フォントや文字色の指定などの見た目の指定、などといった機能がある。

ティム・バーナーズ＝リーによってSGMLを元に開発された。1993年に最初のドラフトが公開され、最初期においてはIETFが、1996年以降はW3Cが、2019年以降はWHATWGが規格の策定、仕様公開を行なっている^{[3][4]}。

特徴 ^[編集]

HTMLは木構造（入子構造）のマークアップ言語であり、形式言語である。「プレーンテキストの文書を要素で括って意味付け」という一般的な説明^[5]は間違っている。「『タグ』と『タグ』で括られたもの全体」が「要素」(element)であり、タグすなわち要素ではない。マークアップ言語としての特徴は、先祖であるSGMLや、兄弟のXMLと共通しているため、以下ではWWWというシステムにおける「ハイパーテキスト記述言語」としての側面についてのみ記述する。

図 1 学習促進処理済 Web ページの例 [3]

Wikipedia では、多くの重要単語がハイパーテキストで表示されている。使用者が閲覧時、意味の分からない語は、リンク先の Web ページを閲覧すればその内容を理解できるよう、ページが構成されている。これは重要語のリンクを辿ることで知識獲得できる点において、提案したいシステムと類似している。自然な反復学習促進の目的でこのような Web ページ作成を目指す。

3.2 課題 2 へのアプローチ

提案アーキテクチャに従い、既存 Web ページに学習促進処理を施したコピーページを作成する。ここでは、以下の課題を解決することを目指す。

- Web ページをブラウズすれば余計な学習なしに記憶できるシステムの提案
- 専門知識を前提としていない Web ページの設計
- 使用者が意図せず反復学習できる環境の提示

検索対象語は ChatGPT を用いて抽出し、は検索データから推定する。ハイパーテキスト化は、学習者に対応して HTML の記述を変更する。このアーキテクチャの静的構造を図 2 に、動的構造を図 3 に示す。詳細は 4 節に後述する。

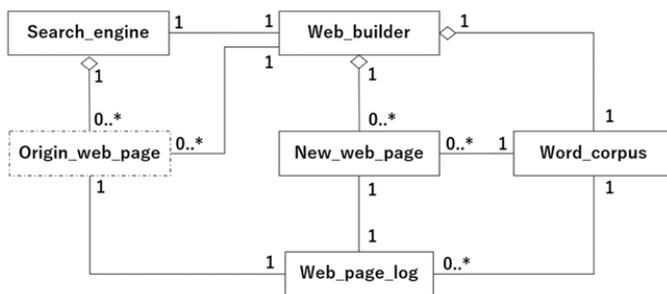


図 2 提案アーキテクチャの静的構造

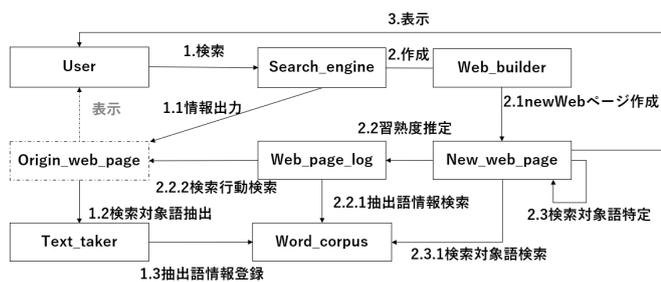


図 3 提案アーキテクチャの動的構造

アーキテクチャの動的振る舞いを図 4 に示す。提案するアーキテクチャにおける、学習促進処理は以下のおこなわれる。

1. 使用者による Web 閲覧
2. 検索情報の収集
3. 使用者習熟度の推定

4. 検索対象語の特定

5. ハイパーテキスト化

使用者の行動から得られる情報をアーキテクチャは取得する。その情報から使用者の習熟度を推定する。検索対象語は、使用者が閲覧した Web ページのテキスト情報から特定される。これは ChatGPT を用いておこなう。この時、難易度と検索対象語を紐づけて出力する。この結果と、使用者習熟度をもとにハイパーテキスト化をおこなない、新たな Web ページが作成される。

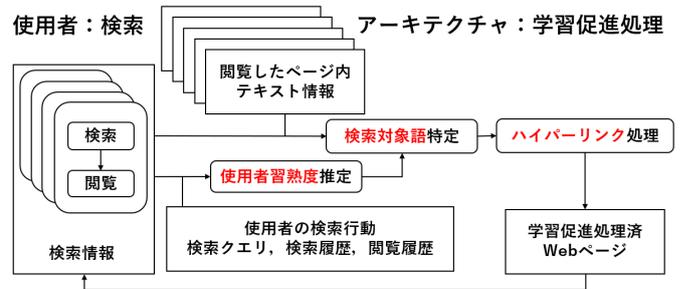


図 4 提案アーキテクチャの動的振る舞い

学習時における使用者の操作は以下の通りである。

1. Web 検索
2. 学習促進処理済ページの閲覧
3. リンク先への遷移
4. 1, 2, 3 の繰り返し

3.3 使用者習熟度の推定

3.3.1 使用者習熟度の定義

使用者が検索時、自然な学習を行えるように使用者習熟度を定義する。この習熟度を用いて、ハイパーテキスト化する語句を確定させる。使用者の疑問箇所にリンクを付与するが、疑問の判断基準は習熟度により変化する。この変化に対応する目的で使用者習熟度を推定する。

例えば、使用者習熟度の分類は 3 段階とする。使用者を習熟度の低い順にビギナー、ベーシック、エキスパートの 3 つに分類する。ビギナーは検索語に対して知識が少ない使用者とする。エキスパートは検索語に対して多くの知識を持ち合わせている使用者とし、ベーシックはそれらの中間である。

ハイパーテキスト化による目標は使用者が習熟度の低い語に対して自然と手が伸びるような状態を目指すことである。ビギナーには多くの語に対して関心が向くようにし、エキスパートには難易度の高い語に対して学習を促すような仕組み作りが必要である。

3.3.2 使用者習熟度の推定手順

使用者習熟度は検索行動から推定する。検索行動とは検索履歴、閲覧履歴である。これらをその専門的な難易度から評価し、習熟度を推定する。検索履歴の判断基準は、検索クエリに入力した語の難易度である。閲覧履歴

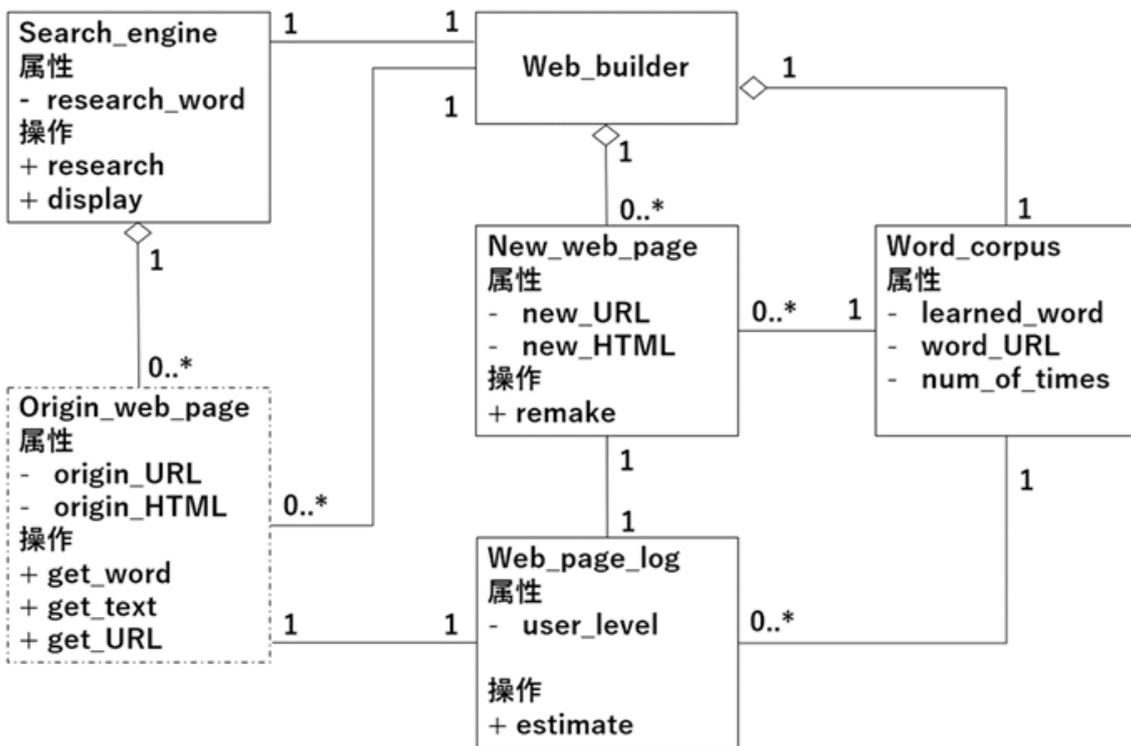


図5 提案アーキテクチャの静的構造 (詳細)

の判断基準は、選択したリンクの語の難易度と遷移回数である。

使用者習熟度の推定は、使用者が検索語入力とリンク選択をするごとに再度おこなう。この繰り返しにより、適切な使用者習熟度をインタラクティブに推定できる。

3.4 検索対象語の抽出

ハイパーリンクを付与するために、テキストから検索対象語を抽出する。抽出語には、その難易度から3段階のラベルをつける。ChatGPTには、以下のプロンプトを入力とした。

```
messages=[
{"role": "user", "context": "以下のテキストの専門用語と概念を抽出して分類してください。専門用語には、3段階で難易度をラベル付けてください。"}
]
```

3.5 Web ページの生成

既存 web ページ情報と使用者習熟度、検索対象語を入力として、新たな Web ページを作成する。データの変換処理を図6に示す。既存 web ページの HTML に学習促進処理をする。これは抽出した検索対象語をハイパーテキスト化するが、使用者習熟度を考慮してどの検索対象語にリンクを付与するか特定した上でおこなう。リンク付与が確定された検索対象語について、HTML の記述を変更し、新たな Web ページの HTML として出力する。

4 設計

図5に示した提案アーキテクチャの静的構造は、Pipe and Filter[4]に基づき、設計した。これは、自然な学習を



図6 Web ページ作成の DFD

支援するシステムの本質はデータ変換処理であるという事実に起因する。これにより、処理に変更があっても独立した Filter を書き換えることで柔軟に対応可能である。

新しく生成する Web ページと生成する際に必要な使用者習熟度、検索対象語を管理出来るように定義した。検索エンジンは既存のものと同様に定義し、Web ページは既存のものとは異なるクラスとした。これは新しく生成した Web ページを閲覧する際に、複数のクラスとの相互作用があることから、シンプルな設計を目指す上で分割した。また、ソフトウェアの柔軟性の観点から、新たな Web ページの作成に関するクラスのインターフェイスとして関数クラスを定義した。

Web builder は新たな Web ページを作成するための関数クラスである。New web page と Word corpus を集約する has-a 関係として定義した。Word corpus の情報は繰り返し利用し書き換える必要があり、所有することが望ましい。New web page クラスも同様に、使用者が操作を行う毎に書き換えるので、関数クラスが所有することで状態管理を行う設計とした。また、Web builder の集

約により柔軟性を高める。外部から隠蔽することで、関数クラス内の構造を隠蔽し、他クラスはインターフェイスのみを使用することができる。所有したクラス内を変更する場合も他クラスへの影響を最小限に抑えることができる。

4.1 検索対象語特定に関する設計

図7に、検索対象語特定に関する静的構造を示す。

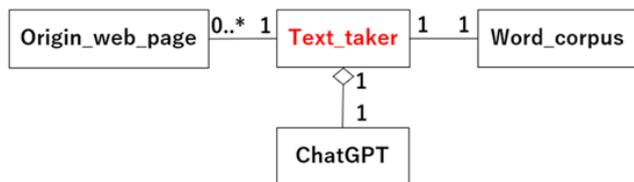


図7 検索対象語特定に関する静的構造

Text taker は関数クラスである。検索対象語の特定に関してカプセル化をすることで外部クラスへの影響を抑える。chatGPTにより、検索対象語の特定をおこなう。その結果を Word corpus が保持する。origin Web page について、HTML の情報を共有する。HTML の情報は ChatGPT の URL Text Extraction によって、ユーザーに表示されているテキスト情報のみを抽出可能になっている。また、抽出後のテキストから ChatGPT は重要語を特定し、検索対象語として Word corpus と情報を共有する。

5 考察

検索者の習熟度向上を目的として提案したアーキテクチャについて定性的に議論する。

5.1 有効性

提案したアーキテクチャの設計について、以下の観点から有効性について議論する。

- 学習効果
- 提案機能の正確性
- 提案機能の妥当性

Web 検索による学習の習熟度は、学習内容に関する問題の正答率を求めることで定量的に判断することが可能である。問題例は、検索対象語についての空欄補充問題が挙げられる。それぞれのユーザーが目にした情報を正しく学習できているか、正答率から判断することができる。

ユーザー習熟度の推定と検索対象語の特定の正確性は重要である。ユーザー習熟度の推定結果は、ユーザーに自然な学習を促進できるかに影響を及ぼす。誤りにより、学習すべき単語をユーザーに正しく提供できない問題が生じる。同様に、検索対象語の特定結果は、促進する学習内容を左右する。どの語が可視化され、どの語についての検索をアーキテクチャが促すかに直結するからである。Web ページ内の重要語句を抜け目なく特定できているかや、特

定した語が検索事項において重要な語といえるかどうか、その正確性は明らかにする必要がある。

ハイパーリンクがユーザーを誘導できれば、妥当性を評価できる。自然な学習を促進する目的で付与したハイパーリンクは、可視化と追加検索の簡易化が利点である。web ページ内の重要語を色づけた注意喚起がユーザーに影響を与えるかは重要である。これらを明らかにすることで、本ソフトウェアが新たに提供する機能が有効であるか評価することができる。

5.2 先行研究との比較

ユーザーの自然な学習を促進することを目指し、設計をおこなった。ユーザーが空欄補充問題を解く等の追加タスクなしで、検索を促すアーキテクチャを提案できた。これは重要語の可視化と追加検索の簡易化の2点が加藤ら [1] と異なる。重要語のハイパーテキスト化により、色がついて強調表示されることは、ユーザーのタスクを増やすことなく学習促進を期待できる。また、促進されたリンク先への遷移は、ユーザーにとって疑問の語句を簡易的に調べることができ、自然な学習が促されることが期待される。このように、自然な反復学習を促進できる点が先行研究と比較して異なる。

6 おわりに

本研究では、Web 検索による語彙学習を支援するアーキテクチャを提案した。既存の web ページに新たなハイパーリンクを追加することで、閲覧者の学習促進を目指した。学習を促す検索対象語は ChatGPT を用いて決定し、ユーザー習熟度を考慮することで、単語の難易度に応じてリンクの付与箇所を確定できるようにした。

今後の課題として、検索対象語をユーザーが十分に学習できるかを明らかにする目的で、本研究の着想は定量的な評価をしなければならない。提案した設計に基づいて実装をし、より複雑なシナリオへの対応するための設計の見直しや、システム間のインタラクションを確認する必要がある。また、アーキテクチャの有効性について定量的な評価をおこない、要素技術についてその妥当性を再検討していきたい。

参考文献

- [1] 加藤 慎融, 柏原 昭博, “Web 調べ学習促進のためのコンテキストウェアな空欄補充問題生成とその評価”, 教育システム情報学会誌, 2024.
- [2] BROWN, Tom, et al. “Language models are few-shot learners.” Advances in neural information processing systems, 2020.
- [3] Wikipedia - HyperText Markup Language, https://ja.wikipedia.org/wiki/HyperText_Markup_Language 最終アクセス 2025.01.17.
- [4] M.Shaw and D.Garlan, Software Architecture - Perspective on an Emerging Discipline, Prentice Hall, 1996.