

ボタンと音声を用いた Web リンク選択インタフェースの提案と評価

秋元 啓孝†

中野 鐵兵†

小林 哲則†

†早稲田大学

1 はじめに

ポインティングデバイスや、十分な数のキーが使用できない状況で、Web ページ上の情報へアクセスする際の、操作性向上を目的とした、ボタンと音声を用いた Web リンク選択インタフェースの提案と評価を行う。

現代の情報社会においては、多くのシステムが Web アプリケーションとして動作しており、モバイル、車載環境からアクセスする機会も増加している。しかし、そのような環境では、ポインティングデバイスやキーボードを使用できない状況も多い。それで数個のボタンのみで操作を行う場合、ページ内のリンクを、1つずつ移動して選択する操作は、とても手間がかかる。汎用的な Web ページで、リンク選択を行うインタフェースとしては、リンクの横についた数字を入力して選択する Hit-a-Hint[1] や、リンクの単語に関連した数個のキーで選択するもの [2] がある。しかし、ともにキーボードの使用が前提で、数個のボタンしか利用できない状況に対応していない。

そこで、あらゆる一般の Web ページのリンクタグを分析し、リンクのグループの間をボタンで移動し、その中のリンクや選択項目を音声で指定するシステムを構築する。具体的には、Firefox 拡張機能と Proxy-Agent[3] による音声認識サーバを組み合わせたシステムで、リンク情報の抽出、リンクのグループ化、VoiceXML と音声認識文法、辞書の生成、ユーザ操作の検出、音声認識、実際のリンク選択、といった処理を行う。

この手法で、実際の Web ページでどの程度、本システムが対応することができるか、また、1つずつのタブ移動で選択する方式に比べ、操作性がどの程度向上するか、検討を行った。

2 実現するインタフェース

提案するインタフェースは、ページ内の全リンクを分割したグループ間を移動し、グループ内のリンクを音声で指定して選択するものである。

操作は、限られたボタンのみが利用可能な環境を想定して、4 方向ボタンと決定ボタンのみが利用可能なコントローラで行う。4 方向ボタンでは、グループ間の移動が可能である。上下ボタンは、HTML 記述順移動とし、左右ボタンは、空間的な左右移動とする。上下・左右とも空間的な移動とせず、上下を HTML 順移動とするのは、ページ内のすべてのリンクに、確実に移動できるようにするためである。空間的な左右移動には、空間的にすぐ右(左)で、感覚的にすぐ移動できるように見えるが、HTML 順では離れているという場合に、移動のストレスを軽減する目的がある。

選択時は、そのグループ上で、決定ボタンを押しながら項目を発話して音声入力する。決定ボタンの短押しで、フォーカスのあるリンクの選択を決定し、実際にリンク先にジャンプする。この操作方法を我々は、Select&Voice[4] と呼ぶ。

リンクのグループは、操作性を考慮し、HTML 内での局所性、構造的な並列性を基準に、グループ数が多くなり過ぎないように作る。具体的には、HTML 内の 2 つのリンク a , b (HTML 記述順) について、共通の祖先から見たタグ構造が一致し、 a とその前のリンクの関係より、 a , b の関係が深い場合に、 a , b 間のすべてのリンクをグループ化する。さらに、これらの処理の

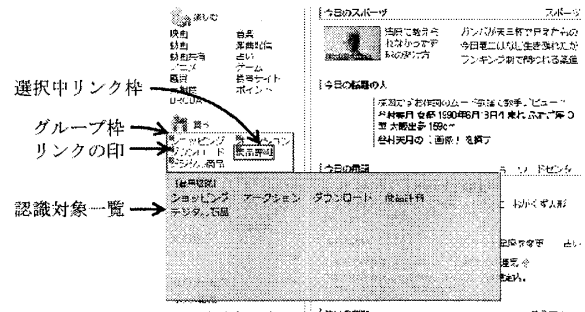


図 1: GUI

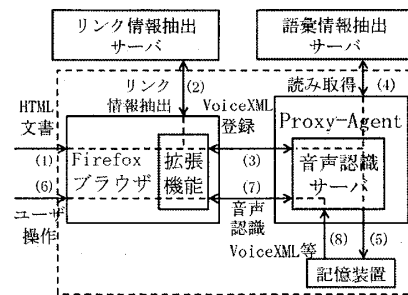


図 2: Web リンク選択システムの構成

みでは、グループ化されない孤立したリンクが生じるので、リンク数が 4 未満のグループは、その前後のうち、共通祖先までたどる要素数が少ない方と結合する。

図 1 は、本システムの音声認識中の GUI である。グループは枠で示され、内部の各リンクには、もしグループ同士の範囲が重なっていても、グループの所属関係が分かりにくくならないよう、印を表示する。グループ内のどのリンクが選択されているか分かるよう、選択中のリンクに、枠を表示する。グループ間の移動時には、移動したことがはっきり示されるよう、枠の周りに表示効果を出す、音声認識動作を分かりやすくするため、音声認識中は、色を赤に変化させ、認識対象テキスト一覧を表示させる。

3 システム構成

本システムは、処理全体を管理する、Firefox ブラウザ上で動作する Firefox 拡張機能、音声認識処理を行う音声認識サーバ (内部)、リンク情報抽出処理を行うリンク情報抽出サーバ (外部) で構成される (図 2)。

3.1 Firefox 拡張機能

Firefox 拡張機能では、ページ読み込時に、HTML から、音声認識による操作を記述する言語である VoiceXML と、音声認識文法を生成し、登録する (図 2 (1) ~ (3))。ユーザ操作時は、その VoiceXML 内インデックスを音声認識サーバに送信し ((6), (7))。認識結果に応じて、実際に選択を行う。VoiceXML は、データの指定が容易な XML 形式であり、広く使われていることから、本システムで採用している。

3.2 リンク情報抽出サーバ

本システムでは、VoiceXML の対話 1 個分に当たる <field> に、リンクグループを対応させ、ユーザ操作時

表 1: リンク抽出・グループ化の性能評価

項目	問題なし	問題あり
(a) ロード時取得インデックスが操作時と一致するか	95	5
(b) 見えるのに選択できないものがないか	69	31
(c) 見えないリンクに印がつく場合がないか	71	29
(d) グループ同士重なる現象がないか	46	54

はこの <field> のインデックスを指定する。<field> 内の文法（選択テキストのリスト）内項目 1つが、リンク 1つとなる。このサーバでは、その項目テキストとして、<a> と の間のテキストを取得する。

リンク情報抽出処理は、あらゆる Web ページに対応させるため、更新が頻繁に必要である。このようにサーバとして独立させることには、それを容易にする目的がある。

3.3 音声認識サーバ

ページ読込時に、VoiceXML と文法を登録する。また、用いられている単語から、語彙情報サーバ [5] により読みを取得し、辞書を生成して登録する (図 2 (3) ~ (5))。操作時は、それで音声認識を行い、結果を返す ((7), (8))。このサーバは、ローカルの Proxy-Agent 環境 [3] で HTTP ベースのサービスを提供することにより動作している。この方式では、クライアント環境に依存せず、自由な音声認識エンジン選択が可能である、という利点がある。

4 リンク抽出・グループ化の性能評価

本システムが、実際の Web ページでどの程度利用可能であるか調べるため、様々な種類の 100 の Web ページを対象に、リンク情報抽出、グループ化を実際に行った。

検討項目と結果を表 1 に示す。問題ありとなる原因の多くは、JavaScript によるページ書換である。(b) については、iframe 等、本システムで対応していないタグ記述、(c) については、スクロールしないと見えないリンクが原因で、問題がある場合が多かった。

インデックスが正しく取得できれば、動作可能であるので、100 ページ中 95 ページで、情報を取得できたリンクに関しては、すべてが選択可能である。但し、操作性面で問題のあるものとして、表示されているのに関わらず選択できないものが 31 ページで、選択対象のリンクが見えないものが 29 ページで、一部にあった。また、半数以上の 54 ページで、複数のグループが重なる場合があった。グループの重なっている部分にリンクがあると、それがどちらのグループに所属するのか、分かりづらくなるという問題が生じる。選択されているグループ内のリンクに印をつけることで、対策を行っているが、さらなる改善が必要である。

5 ユーザ評価実験

本システムの有効性を調べるため、12名のユーザに、Web 上の 20 のリンク (3 ページ) を、本システムと、1 つずつのタブ移動で選択するインタフェース (HTML 順のタブ移動と決定ボタンのみ) で選択してもらい、その所要時間や操作回数を記録し、アンケートにより主観的評価を得る実験を行った。

全ユーザ、全リンクの、移動ボタン押下回数 (押し続けた場合も、1 つ移動につき 1 回と数える)、選択完了までの所要時間の平均を、表 2 に示す。いずれも、本システムの方が、タブ移動に比べて良いことが分かる。また選択時間に関しては、12 人すべてのユーザで、本システムの方が、短かった。

表 2: ユーザ評価実験での操作回数・時間平均

項目	グループ	タブ
移動ボタン使用回数平均	8.69	88.26
リンク選択時間平均	8.09	12.06

表 3: ユーザ評価実験のアンケート結果

項目	評価
(a) 目的リンクに速く到達できる	4.4
(b) 操作のイライラが少ない	4.7
(c) グループ区切りの位置が操作面で適切	4.1
(d) 移動時に枠が光ることで動作が分かりやすい	4.8
(e) 認識時の赤表示・テキスト表示で動作が分かりやすい	4.3

選択式アンケートの結果を表 3 に示す。(a), (b) が、タブ移動との比較 (5 が本システム, 1 がタブ), (c) ~ (e) が、本システムを 5 段階で評価してもらったものである。選択速度や操作時のイライラの、タブ移動との比較、グループ区切り、枠が光る、認識時の表示、といった本システムの評価の項目で、4 点を超えていて、この結果は、本システムは多くの点で、タブ移動に比べ利点が大きく、また、本システムのグループ化や GUI の工夫が、操作性向上に有効である、ということを示している。

本システムの良い点、改善点を自由に記述してもらったアンケートでも、現在の位置が把握しやすい、直観的で分かりやすい、ボタン操作の手間が少なくすむ、左右移動によってストレスがなくなる、といった評価が得られた。一方で、枠が見づらい、発話する文字が分かりにくい、音声認識以外でもリンクに飛べるようにした方が良い、といった改善点も指摘された。

6 まとめ

通常のポインティングデバイスやキーボードが使用できない環境での、Web 上の情報アクセスのユーザビリティ向上のために、一般の Web ページのリンク情報を抽出し、音声によって選択を行うシステムを構築し、実際のページで動作させた。その結果、多くのページでそれが実現でき、従来のタブ移動に比べても利点が多いことが示された。

今後の課題としては、JavaScript で内容が変更されるもの等、より多くの Web ページに対応し、グループ化アルゴリズムや GUI の改善により、操作性が向上するよう、検討を行っていく。さらに、ジェスチャ認識等のインタフェースと統合し、より多様な環境での情報アクセスが実現できるよう、研究を進めていく。

参考文献

- [1] hah.mozdev.org, Hit-a-Hint, <http://hah.mozdev.org/>, Dec, 2008.
- [2] Leo Spalteholz, Kin Fun Li, Nigel Livingston, Foad Hamidi, "KeySurf: A Character Controlled Browser for People with Physical Disabilities," Proceeding of the 17th international conference on World Wide Web 2008, pp. 31-40, Apr, 2008.
- [3] 中野鐵兵, 藤江真也, 小林哲則, "Proxy-Agent を用いた音声認識対応ウェブアプリケーション開発フレームワークの提案と実装" 情報処理学会研究報告, 2008-SLP-70-15, pp. 83-88, 2008.
- [4] 梅本暁, 中野鐵兵, 小林哲則, "GUI とのアナロジーに基づいた音声インタフェースの提案と評価," 日本音響学会秋季研究発表会講演論文集, pp. 63-66, 2007.
- [5] 中野鐵兵, 佐々木浩, 藤江真也, 小林哲則, "集合知を利用した語彙情報の収集・共有・管理システム," 情報処理学会研究報告, 2008-SLP-71-12, pp. 77-84, 2008.

¹<http://www.yahoo.co.jp/>, <http://www.voiceblog.jp/andotowa/>, 他