

手書き文書を用いた知識共有システムの提案とその技術的検証

Proposal of Knowledge Sharing System Which Utilizes Hand-Written Documents and Its Technical Feasibility

加藤公一* 栗原一貴† 大浦弘樹‡
Kimikazu Kato* Kazutaka Kurihara† Hiroki Oura‡

*日本ユニシス株式会社

†産業技術総合研究所

‡ワシントン大学

*Nihon Unisys, Ltd.

† National Institute of Advanced Industrial Science and Technology

‡ University of Washington

要旨

タブレット PC を活用し、手書きデータの収集・検索が簡単にできるシステムを提案する。これにより、「まだ十分に練られていない柔らかいアイデア」を組織内で共有することができるようになると思われる。また、その技術的な実現可能性をプロトタイプにより検証する。

従来知識共有の際によくつかわれている Word や Powerpoint などのコンテンツは「清書」の結果であって、「よく考えた結果」が反映されることが多い。それに対し手書きではまだ十分に固まっていない新しいアイデアが表現されることが多く、本稿で提案するシステムで、従来収集できていなかった初期段階のアイデアが収集できるようになると考える。

そのように新しいアイデアを有効に収集し活用するために、手書き文書を加工せずにそのまま提示する仕組みを前提として考える。また、手書きの持つ即時性・即応性を十分に引き出すために、入力者が後の検索性を特に意識することなく入力できるインターフェースを目標とする。つまり、「紙と鉛筆のようにすらすらと」入力でき、しかもその文書が後に検索できるようにする。

1. はじめに

タブレット PC やタッチスクリーンなどの手書き・手入力のためのインターフェースが普及してきている。携帯電話でいえば Apple の iPhone や HTC の Touch Diamond などがタッチスクリーンを利用している。また、PC 向け OS としては、Windows Vista ではタブレット機能が標準装備であり、その次のバージョンである Windows 7 では、さらにそれに加えてマルチタッチに標準で対応している。

手書きによりテキストを入力しようとする試みは従来からある。PDA などの小型端末では、古くから文字認識技術を利用した手書きテキスト入力が行われてきた。また、Windows Vista では、入力のフロントエンドとして、手書きテキスト入力が標準で用意されている。しかしそれらにおいては、手書き入力は「キーボードの代わり」としての位置づけであり、使い勝手の面からかならずしも手書きのよさを十分に引き出しているとは言いがたい。また、それらにおいて保存対象は電子的テキストであり、手書きによる表現力が十分に生かされていない。

また、絵を入力するためのソフトウェアでは、手書きは従来から活用されている。ペイントツール・ドローツール・オーサリングツールなどでは、自由な曲線を描く手続きが頻発するためタブレット PC 等との相性はよい。また、携帯電話や PDA などの小型端末においては、メモとして手書き入力を画像として利用するためのツールが広く使われている。しかしこれらのツールにおいては、手書きコンテンツは「絵」として保存されるのみであって、それを後で検索できるようにするためには、保存時にタグをつけるなど付加的な作業が必要になる。

本稿では、絵を描くような直感的なインターフェースで手書きを直接受け入れながらも、それをテキストで検索できるようなシステムを提案する。手書きで入力したものをそのまま保存し提示することを前提とし、しかもそのコンテンツが検索可能な仕組みを提案する。手書き入力が本来持つよさを引き出すため、「紙と鉛筆のようにすらすらと」入力できる仕組みを目標とし、できるだけ入力の違和感を排除

し、タグ入力のような付加的な操作をなくすことを目指す。

この様に、手書き入力の操作性と手書きコンテンツの検索性を同時に高めることにより、初期段階のアイデアを収集・活用できるようになると考える。現在さまざまなナレッジマネジメントシステムが提案されているが、その対象となる知識はテキストファイル、Word、Powerpointなどで表現されたものであり、「清書」としての文書に限定されてきた。それに対し、手書きコンテンツを管理対象とすることにより、「まだ固まっていない初期段階のアイデア」も組織内で共有できるようになると考える。

紙と鉛筆のように自然に入力された手書きコンテンツを、タグなどの付加情報を登録者が明示的に入力せずに検索可能にするために、文字認識技術を用いる。しかし、現状の文字認識技術は不完全であり、一定の確率で認識誤りが発生する。しかも操作性を重視し、できるだけ制約のない自由な手書き入力を許すと、従来 Windows Vistaなどで用いられた手法よりもさらに認識精度が低下する。高い操作性と検索性を両立させる手段としては、1)入力時にはさりげなく認識結果(いくつかの候補)を提示するが、その中から正しい結果を選ぶことを強要しない、2)参照時にも認識の誤りを訂正できるようにする、ということ提案する。

提案システムの技術的実現可能性を示すため、プロトタイプを実装した。プロトタイプは、手書きプレゼンプラットフォーム「ことだま」[4,5]の上で開発され、データベース部分はタグベースのファイルサーバシステムである tacoPot[3]の上で実装された。こうして作られたプロトタイプは、オープンソースとしてウェブからダウンロード可能である。

本稿では、手書きを使ったナレッジマネジメントの可能性や、その使いやすさについては仮説にすぎず、まだ十分に検証されていないので注意が必要である。一方、その仮説のもとで、それが技術的に実現可能かについて実証している。手書きを使ったナレッジマネジメントの可能性については2章で仮説を説明する。その仮説のもと、3章ではどのような要件を満たせば使いやすいものになるかについての仮説を立てる。その要件を満たすものの実現可能性について4章でプロトタイプにより検証する。最後に結論を5章で述べる。

2. 手書きを用いた創造的知識共有とフォークソノミー

コンピュータシステムとしてのナレッジマネジメントシステム(KMS)においては、従来から有用な知識を発掘するための検索の仕組みは広く研究されてきたが、入力コストを下げるためのユーザーインターフェースの研究についてはあまり重きが置かれてこなかった。

手書き、音声および身体動作(ジェスチャー)などの人間の直接的作業をインターフェースにより受け入れることは、KMSの入力コスト軽減に大きく役立つと考えられる。例えば、キーボードやマウスは、コンピュータの入力のために専用と考えられた装置であり、それにより入力するためにはその使い方を習得せねばならず、またそのような機械的な入力装置に向き合うには心理コストも高いと思われる。その一方、手書き、音声、身体動作などは、日常生活で対人的コミュニケーションで普通に使われるものであり、それによる伝達能力は特に訓練をしなくても多くの場合身につけている。

しかしそのような人間の直接的動作をコンピュータの直接の入力とするのは、現状では技術的に大きな壁がある。直接動作による入力を後で検索可能にするためには、その入力の中から意味を抽出しなければならない。しかし、コンピュータが人間の直接動作の中から意味を抽出する技術は現状ではまだ不十分であり、その不十分さを補うために、動作に対して制約を設け、いわば人間がコンピュータに擦り寄るというアプローチが一般的に行われてきた。例えば、手書き文字認識では、四角の中に入るように一文字を書く、音声認識では、普段話すよりも明瞭に発音する、などといったことがその例である。また、そのようなアプローチでも認識に誤りが発生する場合(あるいはコンピュータが認識結果に自信がない場合)は、おそらくこれであろうという候補をコンピュータがいくつか示し、その中から正しいものをユーザが選ぶという手法が多く用いられてきた。

しかし、そのような「人間がコンピュータに合わせる」ことを課すと、自由な知的創造が阻害されると考えられる。手書き、音声および身体動作は、リアルタイムな議論では重要な道具であり、その大き

な特徴は即時性と即応性である。つまり、考えをすぐに表現できる、相手の考えにすぐに応答できる、という道具を使ってこそ、自由な知的発想や知的議論が促進される。そのような道具が持つ即時性・即応性は、「コンピュータに合わせる」ことで、大きく失われ、その結果知的創造にも悪影響を与えられる。

そのような問題点を踏まえたうえで、入力者がコンピュータに合わせるのではなく、認識技術の不完全性をフォークソノミー（集合知）で補おうという試みが、近年いくつか出てきている。つまり、入力者ではなく参照者がコンピュータに合わせることで、入力者の知的生産性や創造性を十分に引き出そうという考え方である。例えば、Google Image Labeler[1]というサービスでは、与えられた画像にどれだけ多くのラベルをつけられるかという対戦型ゲームが提供されている。実はこのゲームのプレーヤーはテキストタグを与えることで画像の検索精度の向上に寄与しているのであるが、ゲームのエンターテイメント性により積極的なタグ付けが引き出されている。また、Podcastle[2]というサービスでは、ネットラジオの番組を音声認識した結果とともに保存し、その認識結果のテキストが参照者により編集可能である。これは、ラジオのリスナーが参照した時に、認識の誤りを直してくれることを期待している。人気番組は、認識誤りを訂正する動機づけも高く、検索精度が高まる傾向がある。

そのようなフォークソノミー活用の考え方を、手書きにも応用しようというのが本稿における研究の原点である。特に企業内の知的創造活動においては、手書きの文書は多く発生しているが、その多くは後で参照されることがあまりない。これは検索しづらさに大きな原因があると思われる。近年の検索技術の向上により、電子的にテキストとして保存された文書は、「整理して保存する」という必要性が薄れてきている。また、整理して保存することからの解放を徹底するための研究も近年進んできており、例えば tacoPot[3]はその一例である。そのように、整理するわずらわしさから解放された環境においては、わざわざ手書きのものを「きちんと整理して後で検索可能なように保存する」という動機は以前にもまして小さくなってきている。そのため、整理しなくても手書きを検索できるようにするというのが、本研究の目標である。

また、組織内で手書きを積極的に収集することは、従来では共有が難しかった「初期状態のアイデア」の共有促進につながると考えられる。つまり、従来ファイルサーバや KMS で管理されてきたのは、主に電子的なテキストであり、多くの場合は「よく考えて練られた後の清書」であった。それに対し、手書き文書を積極的に収集することは、清書の前の段階のアイデアを収集することになり、知的創造の初期段階での知識共有を促すことにつながる。

手書き文字にはテキスト以外の非言語的な要素も含まれ、表現力が高い上に、入力に要する時間が短い点が大きな利点である。キーボードとマウスのような装置では、非言語的な要素（図形など）を入力しようとする、その効率は手書きにはかなわない。この様に表現力の高さと即時性を両立していることにより、初期段階のアイデアの議論においては手書きが多用されるものと思われる。

以上のことをまとめ、目指すシステムを概念を図1にまとめる。文字認識の不完全な点をフォークソノミーで補い、その結果、手書きで書かれた文書群から容易に検索できるようなものを目指している。この結果、手書きで書かれた初期段階のアイデアを組織内で共有できると考えている。

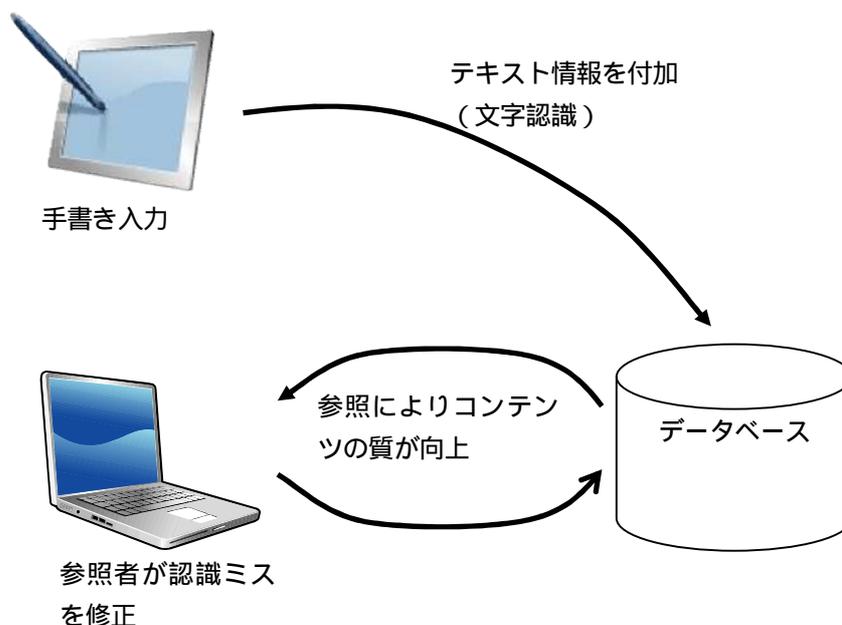


図1 機能概念図

3. 操作性と検索性のトレードオフ

ユーザが紙と鉛筆のように自由に違和感なく文字を書くということと、そうしてできた手書き文書が容易に検索できるかどうかはトレードオフの関係にある。手書き文書をテキスト検索するためには、文字認識技術の利用が必要になるが、技術的な限界により、入力への制約が緩ければ緩いほど文字認識の精度は低下する。そのため、例えば Windows Vista の手書き入力ツール(図2参照)では、一文字を指定された四角の中に書くよう強要するという手段がとられているが、それでは違和感なく自由に書くということが達成できない。



図2 Windows Vista の手書き入力ツール

このようなトレードオフの関係にあるものの双方をできるだけ尊重する手段として、本研究では以下のような方針をとった。

- とにかくどこにでも自由に「紙と鉛筆のように」書けるということを最大限に尊重し、そのために認識精度が犠牲になるのは容認する。
- 時系列的なストローク情報、つまり、いつ書かれたか、いつ書く手が止まったか、という情報を利用することで認識精度を高める。
- 文字認識の結果(いくつかの候補)をユーザに提示することで、認識誤りの修正を促す。しかし、違和感なく自由という点を尊重するため、誤りの修正は任意であり、認識結果の提示はあまり目立たぬようにする。こうすることにより、入力作業の不自然な中断を防止する。また、こうすることで入力者の動機づけの強さにより検索単語の選別が行われる。つまり、この文書はこの単語で検索されてほしいと入力者が強く願うときには、積極的に認識誤りを訂正するであろうと思われる。
- 認識の誤り(認識結果に自信のないもの)も含めて、積極的に保存して活用する。このとき、手書

きストロークと認識結果を関連付けて保存し、参照者にはその関係性を示すことにより、誤りの修正を促す。このようにフォークソノミーを活用することで、重要な文書ほど検索精度（適切に検索される可能性）が高まると予想される。

自由に書けることを最大限に尊重するために技術的課題が多く発生してしまった印象を受けるかもしれないが、これは知的創造を支援するという本来の目標に照らし合わせると妥当な制約である。また、このような技術的課題の発生により、その実現可能性が自明ではなくなった。プロトタイプ実装は、その技術的実現可能性の確認のためのものであり、次章で詳しく説明する。

4. プロトタイプ実装

自由に違和感なく「紙と鉛筆のように」書けるということを最大限に尊重しつつ、そうしてできた手書きコンテンツが検索可能な仕組みを構築できるかという点について、実装を行い実現可能性を確認した。また、認識結果を参照者に提示し、誤認識の訂正を促すことが、適切なユーザビリティのもとに可能かどうかを確認した。

開発の基盤としては、手書きプレゼンプラットフォーム「ことだま」[4,5]を用いた。ことだまはタブレット PC 向けにチューニングされたユーザインターフェースを持ち、ユーザビリティの面での利点が大きいと考えたからである。また、ソースコードが公開されており、プラグイン方式により追加機能が開発しやすいことも理由である。

もともと、ことだまは手書き入力によりできた文書を通常のファイルとして保存するものであったが、その保存先として tacoPot[3]を指定できるようにした。tacoPot はタグ指向のファイルサーバであり、文書の検索性の検証に都合がよいと考えたからである。

文字認識についてはペンが一定時間止まったときをトリガとして、自動的に実行されその付加情報がドキュメントに付加されることとした。認識エンジンは Microsoft Tablet API に付属のものを使った。この API に対しては、ペンストロークの時系列情報を呼出側で意識した上で入力を与えた。つまり直近の入力結果のみを認識の対象とすることで文字認識の精度を上げた。

ペンが一定時間止まった時を文字認識開始のトリガとしたのは、ユーザビリティ上の理由による。つまり、ユーザが明示的に何らかのアクションで「文字認識しろ」と命令するようなユーザインターフェースでは不自然な動作が必要になり、「紙と鉛筆」のような自然さが損なわれる。ペンが止まるという事象は、紙と鉛筆で文書を書いているときにも自然に発生するのでユーザが意図的に認識開始を指示する必要はなく、さらに自然な切れ目を認識の切れ目とすることができる。

認識結果は画面の下方に目立たないように表示することとした。このとき、候補を確からしい順に表示するが、必ずしも候補の中から正しいものを選択する必要はなく、認識結果表示を無視して書き進めることができるようにした。こうすることで、「この文書はこういう単語で検索されて欲しい」という強い動機付けを持ったユーザは簡単な操作でその要求を満たすことができるが、その一方で入力を中断されてほしくないという要求も満たすことができる。

文字入力操作のシナリオを実際の実行画面を使ってまとめると、図3のようになる。このように、文字認識の結果は画面下方に目立たないように表示されるので、無視して書きつづけることが可能である。気になるときに限って、認識誤りを修正すればよい。

認識結果の表示は、対応する矩形領域と同時に表示する。参照者が文字認識の誤りを直したいときは、対応する領域をクリックし、キーボードにより修正する。こうすることで、利用者に面倒な操作を課さずに、フォークソノミーを有効活用できる。

全体の流れのシナリオは図4のようになる。これは、一部は正しく認識できたが、一部は認識を誤った例である。図4において、「文字」の部分は正しく認識できているので、後に「文字」をキーとして検索することは可能である。そうしてこの文書を見つけ出した参照者が、「認識」の部分の認識誤り（「言・忍・識」と間違えている）に気づき、訂正することが可能になっている。つまり、最初の文書作成の段階では文字認識の結果が完全ではなくても、重要なキーワードさえ認識されていれば後で検索可能であ

る。あとは、検索した参照者が自分の考えに基づいて誤認識を訂正すればよく、直したい人が直せばよい。こうすることで、参照回数が多く重要度の高い文書については、誤認識の訂正が積極的に行われるであろうと思われる。

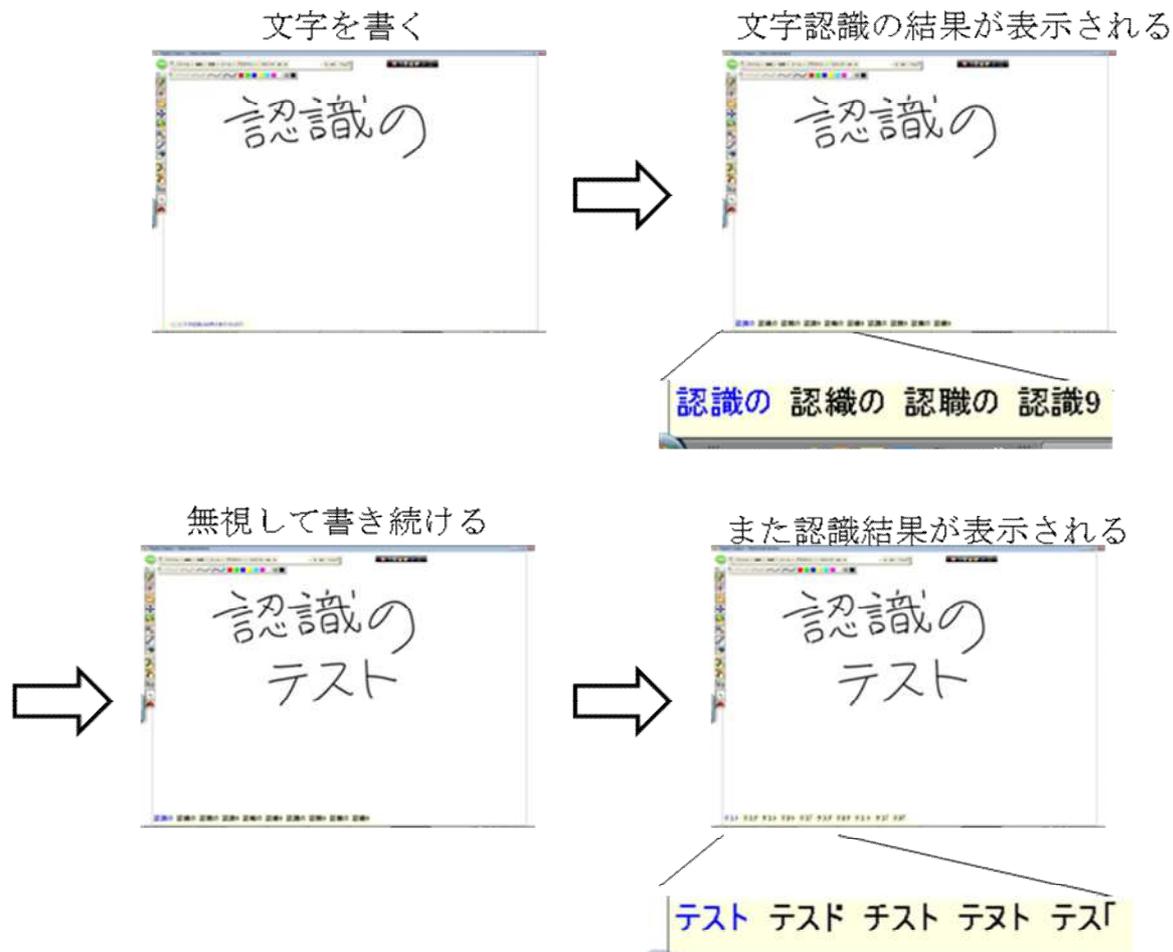


図3 文字入力操作シナリオ

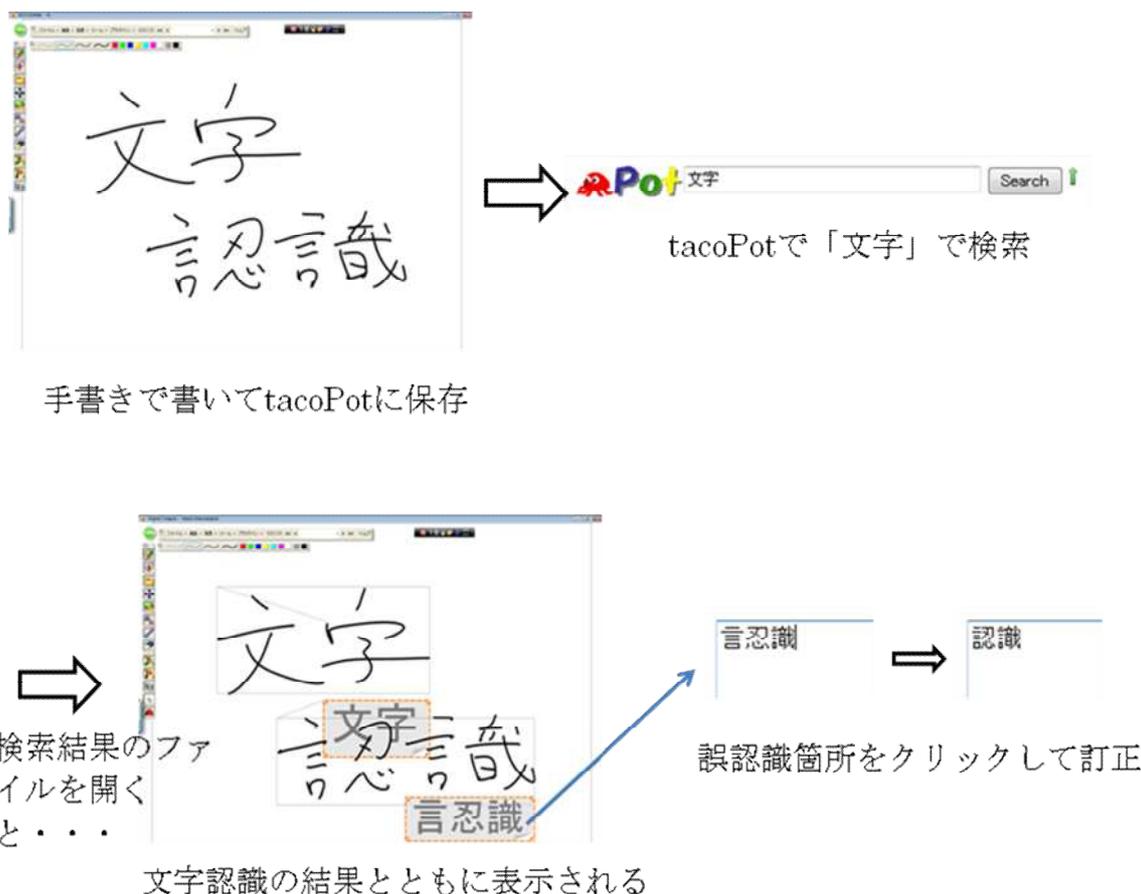


図4 保存・検索・修正の一連の操作シナリオ

5. 結論と今後の課題

本来の手書きの良さを損なうことなく、自由に違和感なく書けるということと、そうしてできた文書を電子的に保存してテキスト検索可能にするということは、ある程度両立できることがプロトタイプによって示された。このことにより、例えばホワイトボードの前で自由に書きながらプレーストリーミングした結果が、何の加工もせずに後で検索できるような仕組みが構築できるようになる。このことは組織の知的創造性の促進に役立つと考えられる。

しかし、プロトタイプによって示されたのは技術的観点からのみの実現可能性であることに注意が必要である。こういう仕組みが一般のユーザにとって利用しやすいかというユーザビリティの評価や、実際に知的創造能力の向上に役立つかという有効性の評価については、さらなる研究が必要である。

また、プロトタイプの結果、技術面についても課題がいくつか見えてきた。特に重要だと思われるのは、以下の点である。

- 文字認識トリガの問題：現状では「手が止まった時」に文字認識エンジンが走ることになっているが、それだけでは不十分であると考えられる。例えば、今まで書いていた場所と大きく離れた場所に書き始めた場合をトリガにするなどの策が考えられるが、「大きく離れた」という基準をどうするかという問題が残る。また、例えば漢字を一時的に度忘れした場合のように、手が止まっても文の途中であって、認識をしないほうが妥当な場合もある。
- 参照者への認識結果提示のユーザビリティ：「文字認識結果表示」ボタンを明示的に押さないと認識結果が表示されないような仕様では、そもそもそのボタンに気付かないことにより誤認識の訂正がなされないことが多くなる。その一方、ファイルを開いたときにすぐに認識結果が提示されていると、参照者にとって必要でない情報が多くなり見づらくなる。つまりフォークソノミーの有効活用と、参照者のユーザビリティの両立が課題である。

これらについてもさらなる研究が必要である。

参考文献

- [1] Google Image Labeler: <http://images.google.com/imagelabeler/>
- [2] Podcastle: <http://podcastle.jp/>
- [3] tacoPot: <http://dev.tyzoh.jp/trac/folksonomy/>
- [4] 栗原, 伊藤, 五十嵐, 編集と発表を電子ペンで統一的行うプレゼンテーションツールとその教育現場への応用. コンピュータソフトウェア, Vol.23, No.4, pp.14-25
- [5] ことだまプロジェクト: <http://dev.tyzoh.jp/trac/kotodama/>