

Gorotte:情報の消費者ではなく生産者になるための試み

Gorotte:Attempt to become the content creator , not just the content consumer

大坪 五郎*

Summary. インターネット上のサービスを利用して蓄積した情報を閲覧し、自由に組み合わせることによって新たなアウトプットにつなげるためのシステム Gorotte を開発した。ユーザは表示された文章からのキーワード選択、あるいは対象文書を指定しての類似文書検索を行うことにより、興味を持ったトピックに関連する文章を画面上で並べて表示することができる。またシステムは、ユーザが選択した単語から自動的に短文を生成して提示する。これらの情報を閲覧することにより、ユーザは保存した情報から新たなアウトプットにつながる気づきを得ることができる。開発したシステムを一定期間使用した結果、情報をより広く収集しそれらについて深く考えるよう、情報との付き合い方が変化した。

1 はじめに

インターネット及び携帯情報機器の普及により、時と場所を選ばず大量の情報を閲覧することが可能となった。また閲覧した情報の中から、興味を引いたものを保存し、検索して取り出すことも可能になっている。

こうした機能は便利ではあるが、それがもたらす負の面についても考える必要がある。情報がいつでもインターネット経由で取り出せると知っている場合、ユーザは情報がどこにあるかだけを記憶し情報そのものを記憶しなくなるという研究結果が示されている [6]。

この結果が意味するところについては様々な議論があり得る。「そもそも情報を記憶する必要はなく、状況に応じて検索して取り出せばよい」という考え方もあるだろう。記憶は変化しやすいこと [11] が知られており、そうしたあやふやな記憶に頼るより、情報をそのままの形で保存しておくほうが望ましいという指摘もある。しかしここで問題としたいのは、情報を保存したことで安心してしまっていないか、という点である。

我々の思考方法は、使用するメディアによって影響を受ける [7]。インターネット及び携帯情報機器というメディアを用いることにより、自分が「保存」した情報を、「知識」として持っているとは錯覚し、それらについて深く思考することを放棄してはいないだろうか。

記憶し、情報を反芻し、新しい物を作り出す能力こそが機械と人間を区別している重要な部分であり、そうした「知を必要とする作業をコンピュータにゆだねまいとする自覚と勇気を持つこと」こそが人間性を失う危機に対処するために重要である、とワイ

ゼンバウムは述べている [8]。同様に情報の消費だけではなく、知的生産が重要であるという指摘がなされたのは 1969 年であった [14]。その後インターネットの普及により情報へのアクセス、及び保存は格段に容易に行えるようになった。すなわち情報のインプットは飛躍的に増大したのだが、それをアウトプット-知的生産に結びつけるための手段はあまり進歩したとは言えない。情報の流入が増大することに満足してしまうのではなく、それをアウトプットにつなげることの重要性は近年とみに高まっているのではないだろうか。

こうした考察により、本研究では閲覧、保存した情報を用いて新たな情報を生成することを目的に、以下の機能を持つシステムを開発した。

- ユーザが自分で保存した情報を振り返る機能。
- 保存した情報の柔軟な組み合わせ、閲覧を可能にすることにより新たな文章作成を支援する機能。

各機能の必要性について説明する。まず自分が収集した情報を振り返る機能を持つことで、情報を保存したことで記憶したような気分になるのではなく、それを反芻し、考察することで記憶の中に取り込み、脳内でより有機的に利用することが可能になることが期待できる。

次にそれらの情報を自由に並べ組み合わせることにより、情報間の関係性についての思考が促され、結果として新たなアウトプットにつなげることができると考えた [15]。

以下、こうした考え方に基づき開発したシステム Gorotte について説明する。

2 Gorotteの機能

2.1 想定する使用シナリオ

Gorotteは以下のような状況で使用されることを想定している。

- ユーザは日常生活において、様々な情報にアクセスし、あるいはそれをきっかけとして思考する。そうした情報を Evernote (<http://www.evernote.com/>) を用いて保存する。
- サーバ側の Gorotte は Evernote と情報を同期し、必要な処理を行う。
- iPad 上の Gorotte はサーバから情報をダウンロードする。
- ユーザは iPad 上で保存した情報を自由に閲覧し、振り返る。関連する情報同士を並べ、集めることにより新たに創造する文章の核となる部分を創り出す。
- 創りだされた情報を Evernote に送信、記録する。ユーザはそれを元にエディタなどを用い文章を記述する。

上記のシナリオを想定した理由について説明する。Gorotte を用いて文章を作成する方法論として、G.M. ワインバーグの「自然石構築法」[3] をベースとする。「自然石構築法」による文書作成は、まず世の中で見聞きするすべてのことから文章の題材を集めることからスタートする。(この題材を「自然石」と呼んでいる) 集めた題材-自然石を用い、あたかも自然石で壁を構築するように組み合わせ文章を構築するのが提唱されている手法である。

ここでユーザが収集した自然石の保存には、既存の Evernote を用いる。また自然石をある程度選別し、並べた以降の文章作成には既存のエディタが十分な機能を持っていると考える。Gorotte が支援するのは、その間、すなわち収集した自然石を反芻し、内容について考え、自由に並べる事により、そこから文章の核となる部分を作り出すプロセスである。

2.2 Gorotteのインタフェース

Gorotte クライアントの画面及び操作の一例を図1に示す。起動すると下部にあるツールバーを除いては真っ白な画面が表示される。画面上をタップすると Evernote に保存した文章がランダムに選択され表示される。この段階で表示されるのは、文章のみであり、Evernote のノートに含まれる画像、クリップした日付、ノートのタイトルなどはあえて表示していない。これは文章が他の情報に埋もれ、内容以上に立派に見えてしまうことを防ぐためである。いわば保存した文章を「裸の状態」で提示することに

より文章自体が持つ意味を明らかにすることを狙っている。

ここからユーザはいくつかの方法で文章を閲覧することができる。表示された文章に興味をひかれないう場合、再度タップすれば異なる文章が表示される。表示された文章自体をさらに詳しく読みたい場合には、文章上でピンチ (拡大) 操作する。すると別ウィンドウで文章の題名、クリップした日時などと共に文章全体が表示される。この画面において、特に関心を引いた部分をドラッグで選択することにより、画面に表示される部分を変更することができる。

ユーザが単語を選択する際の参考とするため、複数の文章で用いられている特定の品詞語 (名詞、動詞、形容詞) は他の文字と異なる色で表示されている。文章中の任意の単語をタップすると (図1:番号1) 選択されたキーワードが画面下部に移動するとともに、共通のキーワードを持つ別の文章が表示される (図1:番号2)。キーワードが二つ選択された場合、Gorotte は後述する短文作成アルゴリズムを用いて、選択された二つのキーワードをつなぐ短文を作成する (図1:番号3,4)。これは、生成された短文をユーザが見てインスピレーションを得るきっかけとするためである。

表示された文章が悪くはないが、もっと他の文章がないかと考えた場合には、ユーザは該当文章をスワイプする。左または右にスワイプした場合は、対象となった文章と類似した傾向を持つ文章を、上または下にスワイプした場合には、表示されている他の文章と関連を持ちつつスワイプの対象となった文章とは異なるものを表示する。この際用いられる文書選択のアルゴリズムについては後述する。

表示された文章に興味を惹かれ、保持しておきたい場合、文章を長押しすることでロックすることができる。ロックされた文章は画面に表示され続ける。また他の文章が表示される場所を確保したい場合には文章をピンチ (縮小) 操作することにより、一行のみ表示にすることができる。さらにピンチ (縮小) すると文章は画面上から削除される。

このような操作を繰り返し、画面上に十分な量の文章が表示されたと判断した場合は、その文章群の情報及び Gorotte が生成した短文をメールとして送信し、Evernote 上のノートとすることができる。送信された情報を元に実際に文章を作成する作業には既存のワードプロセッサを用いることを想定している。

上述のように、Gorotte はあくまでも自分が保存した情報の閲覧のみを行うようにしており、外部の Web 情報などは意図的に表示していない。これは自分がどのような情報を重要と考えているかをありのまま表示し、それらを客観的な視点から眺められるようにするためである。

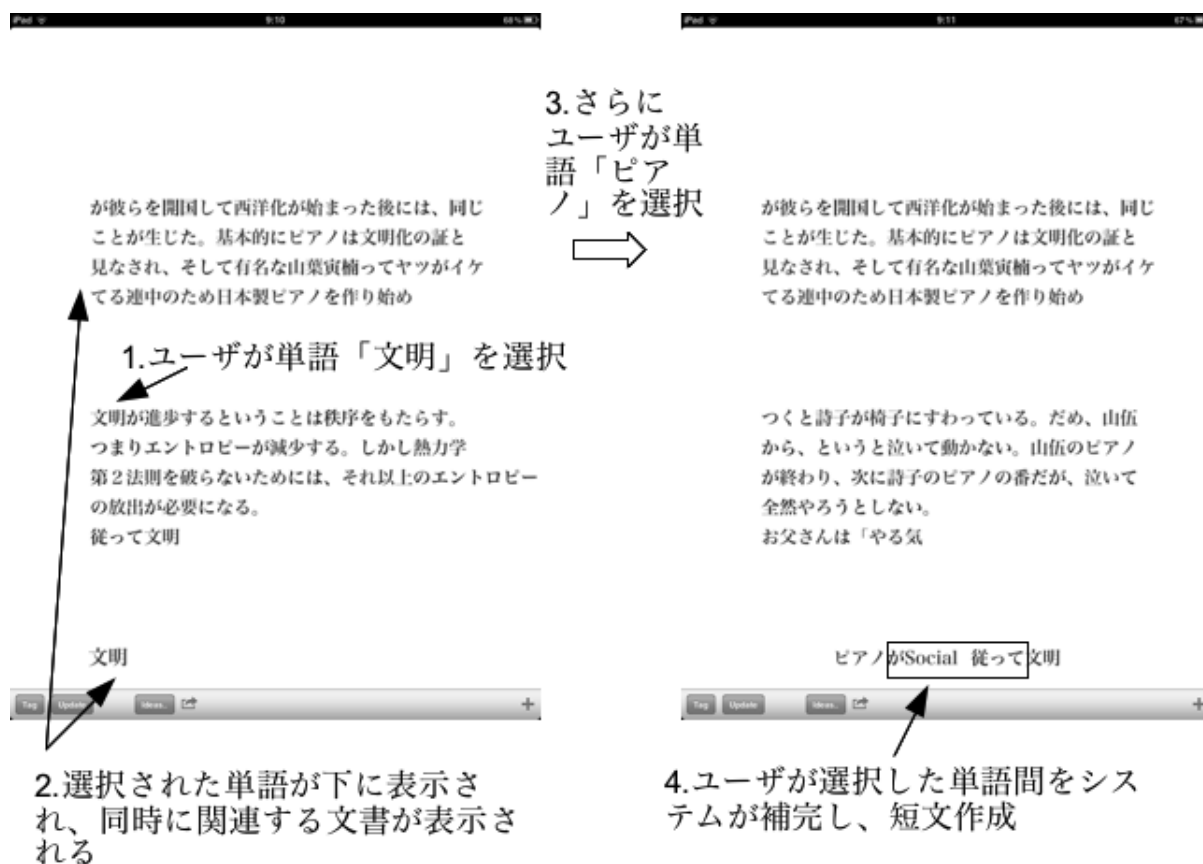


図 1. Gorotte の画面及び操作例

2.3 文章選択アルゴリズム

保存された多数の情報から新たな文章の核となるものを集めるためには「集中と発散」が必要であると考えた。まずユーザは特定の目的を持たずに情報の振り返りを始める。この場合には、次に表示される情報が、現在表示されている情報と関連性を持っている必要はない。ランダム性を持って、幅広い分野の情報を網羅することができるのが好ましい。すなわち「発散」することが求められる。閲覧中に「これは面白いタネかもしれない」というものに出逢えば、その文章に関連の深い他の文章を探し始める。すなわち「集中」である。こうした過程を経て提示された情報からなんらインスピレーションが得られない場合には、異なった情報を閲覧したくなる。すなわち再度「発散」が必要になる。このように「集中」と「発散」を容易に切り替えて実施できることが必要となる。

こうした機能を実現するため、各文章を TFIDF を用いて抽出した特徴語を用いてベクトル化し、Rocchio のアルゴリズム [5] を用いて文書検索を行うこととした。このアルゴリズムでは、適合文書、不適合文書両方を考慮した上で文書検索を行うことが可能である。前章で述べた文書検索のための操作が行

われた際、Gorotte は以下のような処理を行なっている。

- 文章中からの単語選択: 選択された単語 (1-2 個) を特徴ベクトルに変換し、それを適合文書として扱い文書検索を行う。
- 表示された文書を左または右にスワイプ: 表示されている全ての文書を適合文書として文書を検索する。
- 表示された文書を上または下にスワイプ: スワイプ対象となった文書のみを不適合文書、他に表示されている文章を適合文書として文書検索を行う。

上述の操作のうち、上下のスワイプを他に複数の文章が表示されている状態で行うと「集中」、スワイプ対象となる文章しか表示されていない状態で行うと、不適合文書のみ指定したこととなり「発散」をすることができる。

この際ユーザの操作によっては検索結果が存在しない場合がある。通常の文書検索であれば「検索結果なし」とするが、ユーザが発想を得るためには、指定された条件に文書が適合しない場合でも、とにかく結果を返すことが必要であると考えた。そのた

めこうした場合には、直近で表示したものを除いた文書集合からランダムに結果を返すようにしている。

2.4 短文生成アルゴリズム

Gorotte では、ユーザが選択した二つの単語の間を結ぶ短文を生成し、それを提示することによってユーザの発想を促すことを狙っている。この際、短文を生成するアルゴリズムとして、単語単位の n-gram (マルコフモデル) [10] を用いた。具体的には以下のような手順で短文作成を行う。

1. 事前準備として、文章中で単語がどのように連続しているかの情報を作成する。
2. 選択された単語を先頭の単語、その前に選択された単語を文末の単語とする。
3. 個々の文章中における単語の接続情報を参照し、先頭の単語からは後ろ方向に、文末の単語からは前方向にそれぞれ N 単語の列 (以下「単語列」と称する) を作成する。この時点で生成された単語列は、各文章中で実際に使われているものとなる。
4. 単語列の端となった単語のみに着目し、その単語の前後にどのような単語が使われているか、対象とする文書群全体から抽出したデータの中からランダムに選択し、単語列に加える。
5. 先頭の単語および文末の単語から生成された単語列の終端に、共通の単語があるかチェックする。あればその単語を用いて単語列を結合し、処理結果とする。ない場合には、上記手順 2-4 を繰り返す。

基本的なアルゴリズムは上記の通りであるが、開発途中のアルゴリズムを使用した結果、ユーザの思考を途切れさせないため、短い時間で応答を返すことが重要であることが判明した。そのため上記手順 2-4 の繰り返しは一定回数で打ち切り、その時点で生成されていた単語列を連結して結果として表示している。また単語列の長さ N を 2-4 の範囲で変化させたところ、N が大きいほうが意味が通る文章になりやすいが、生成された文章が長くなりすぎるものが判明した。そのため $N=2$ を用いている。

本アルゴリズムを用いて生成された短文の例を以下に示す。文として意味をなさなくても、これらの例はいずれもユーザ (著者) にとって考える元となるものであった。

音楽は裏切りのリズム
放送がずれると目から発想
流刑終了してね?じゃあお休み
集合という理由な能力
ソニーがプリントアウトし部下がポケモン

3 実装システム

Gorotte はサーバサイドアプリとクライアントアプリから構成される。サーバサイドアプリは、Google App Engine (GAE :

<http://code.google.com/appengine/>) 上で動作し、Java で記述されている。サーバサイドアプリの主な役割を以下に示す。

- Evernote システムにアクセスし、ノート of データを同期する。この際ユーザが指定したタグを持つノートのみを対象とすることにより、対象データにノイズが交じることを防ぐことができる。
- ノートからメインとなる文章をある程度識別して切り出す。このアルゴリズムは、extract-content.rb [13] のアルゴリズムを参考に Java で実装している。
- 文章を形態素解析し、結果をクライアントからの要求により送信する。形態素解析エンジンとしては、GAE 上で動作可能な Igo (<https://github.com/sile/igo-gae>) を用いている。

クライアントアプリは、iPad アプリとして構築され Objective-C を用いて記述されている。

4 運用結果

実装したシステムを改良しつつ、実際に蓄積した Evernote のノートを閲覧し、組み合わせによって文章の発想を得ることを試みている。その結果について記述する。

文書数は 1300 程度からスタートし、一日 5-6 件のペースで増加している。対象とする文章の種類を以下に列挙する。

- Web 閲覧中に保存した Web ページの情報
- 頭に浮かんだアイディアのメモ
- Twitter での自分の発言
- ブログに記載した内容

これらの情報を閲覧し、実際に一つのトピックに関して集めた文書群の例を以下に示す。

- 会話する際に、相手の話を勝手に要約してはいけない
- 話しを聞く達人だった河合隼雄氏の会話方法。相手の話を要約しなかった。
- 女性との会話で注意すべき点について。相手を全肯定し、決して否定しない。
- 「なぜ」は、やる気と成長を止める質問
- 「なぜ」を 5 回繰り返すトヨタの方法論

- 生成された短文：「聞くがアーティスティックで引用が爆発」

最初は特に目的なく閲覧を始め、「相手と会話するときの方法論」について興味を持ち、キーワード選択、スワイプによる文書閲覧を経てこのような文書群を作成した。作成された結果を見ると、個々の文書における主張が微妙にずれていたり、対立していることに気がつく。個々の文章を保存する時は、それぞれの内容に意義を認めていたわけだが、このように並べてみると自分が相矛盾する主張に同意していたのではないかという疑問が生まれ、それぞれの内容についてより深く考えることになった。そこから何らかの結論が生まれる場合にはブログなどの形で新たな情報として発信する。

本システムを用いこのようなサイクルを繰り返しているうちに、情報の収集、閲覧、活用の方法が変化した。内容を以下に列挙する。

- Evernote に情報を記録する際、Web ページを丸ごと記録するのではなく、自分で「どこが重要か」考えた上で必要と思われる部分だけを記録するようになった。これは Web ページをそのまま記録した場合、不要な部分が多く表示され後から閲覧するのに不便であることを実感したためである。
- 上記内容とは矛盾する傾向だが、文章全体が興味深く保存する対象を限定できない場合には文章全体を Evernote に記録するようになった。そうした文章を本システムを用いて閲覧すると、自分が当初クリップするきっかけとなった以外の点について興味深さを発見することがあった。こうした用途を考えた場合、現状用いている TFIDF は文書全体から特徴語を抽出しているため必ずしも望ましくない、文書内に複数存在するトピックを抽出できる LDA[2] を用いることが有効であると考えられるがこれは今後の検討課題としたい。
- 収集した情報を閲覧していて、それらが特定のトピック、立場に偏っていてつまらない事に気がついた。また表示された文面に興味を惹かれた文章の詳細を参照した際、それが自分が高く評価していない人物の言葉であることを知り驚いた。これらの経験を経て、それまでは自分と同意見の情報ばかり保存していたのを改め、少しでも感情を刺激された文章は、その内容に同意するか否か、あるいは発信した人物の好悪に関わらず保存するようにした。またそうした異なる立場を考えた上で情報をアウトプットするようになった。

5 関連研究

情報を創造するためのインタラクション支援としては、中小路らによる ART システムの研究がある[9]。ill-defined なデザイン作業を対象としている点、および文章をエレメントとして自由に組み上げることで文章を構成する点などが本研究と類似する。相違点としては、ART システムでは最初からユーザが文章を書くことを目的としているのに対し、本研究では収集した情報を自由に組み合わせることにより、文章を作り上げるきっかけを得ることにフォーカスしている点、及び文章エレメントから通常の文章を編集する部分については対象としていない点を挙げることができる。すなわち本研究は ART が目指しているものよりも、前段階の作業を対象としているといえる。Evernote Sticky[1] は Evernote に保存した文章を PC デスクトップ上の付箋という形で表示するプログラムである。文章を自由に配置し、発想のきっかけとする用途に用いることができる点が本研究と共通する。しかし Evernote Sticky は選択した文書と類似の文章を検索する機能等は持っていない。Web 上のコンテンツを蓄積・分類・検索・管理する手段としてのソーシャルブックマークに関する研究は数多くなされているが、そのほとんどは効率よく情報を集取することにフォーカスしており集めた情報をどう扱うべきか、という観点を持つものは少ない。近藤 [12] らは、都市建設シミュレーションゲームの機能を取り入れることにより、ソーシャルブックマークの整理を動機づけることを狙っている。情報を収集するだけでなく、その後継続的に整理、見直しすることを目的としている点が本研究と共通する。中村は収集した膨大な写真データを死蔵させることなく有効活用するため、日常的に接するカレンダーをインタフェースとして閲覧、検索することを可能とした Calendar for Everything[4] を提案している。対象としているメディアは異なるが、根底にある情報量爆発に関する問題意識が本研究と共通する。

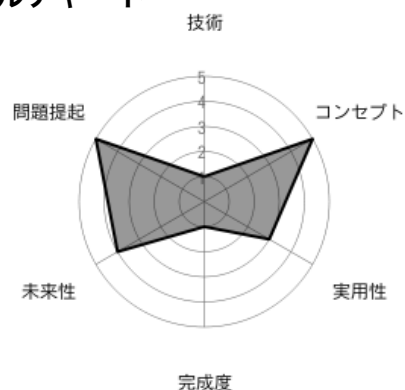
6 まとめ

Evernote を用いて保存した文章を振り返ることを可能にし、情報アウトプットのきっかけを得るためのシステム Gorotte を開発した。本システムは「インターネットの普及によりユーザが情報の受動的かつ浅い消費者」と墮しているのではないか、という問題意識から作成したものである。しかしながら、そうした問題に対する回答は他にも様々な提案があるだろうし、あるべきだと信じている。本論文がそうしたシステムの必要性に関する議論の一つのステップとなることを期待している。

参考文献

- [1] Evernote Sticky: Evernote用デスクトップ常駐付箋紙. <http://sourceforge.jp/projects/evernote-sticky/>, 2011.
- [2] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- [3] G.M. ワインバーグ. ワインバーグの文章読本. 翔泳社, 2007.
- [4] S. Nakamura. Calendar for Everything: Browsing and Finding Cross-Media Personal Contents by Using Calendar Interface. In *Proceedings of the International Conference on Informatics Education and Research for Knowledge-Circulating Society (icks 2008)*, pp. 157–160, Washington, DC, USA, 2008. IEEE Computer Society.
- [5] J. Rocchio. Relevance feedback in information retrieval. *The Smart system - experiments in automatic document processing*, pp. 313–323, 1971.
- [6] B. Sparrow. Study Finds That Memory Works Differently in the Age of Google. <http://news.columbia.edu/research/2490>, 2011.
- [7] ニコラス・G・カー. ネット・バカ, 第7章, pp. 197–198. 青土社, 2010.
- [8] ニコラス・G・カー. ネット・バカ, 第10章, p. 285. 青土社, 2010.
- [9] 中小路 久美代, 山本 恭裕. 創造的情報創出のためのナレッジインタラクションデザイン (AI フロントティア論文). *人工知能学会誌*, 19(2):235–246, 2004-03-01.
- [10] 高橋瑞希, R. Rzepka, 荒木健治. Web 検索と単語 n-gram モデルを用いた文生成手法の性能評価. 言語処理学会第 16 回年次大会発表論文集, 2010-3.
- [11] 小松 伸一, 太田 信夫. 記憶研究の現状と今後. *教育心理学年報*, 38:155–168, 1999-03-30.
- [12] 近藤 拓也, 武田 信也, 品川徳秀. ブックマーク整理活動の持続をめざしたゲーム的 SBM インタフェース. 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM 2010) 論文集 (電子出版), 2010.
- [13] 中谷 秀洋. Web ページの本文抽出. <http://labs.cybozu.co.jp/blog/nakatani/2007/09/2007-09-12>.
- [14] 梅棹 忠夫. 知的生産の技術, pp. 8–20. 岩波新書, 1969.
- [15] 梅棹 忠夫. 知的生産の技術, 第 3 章, pp. 58–59. 岩波新書, 1969.
- [16] 荒井 裕彦. 学術的ロボット研究の問題点について. <http://staff.aist.go.jp/h.arai/robotics/jra03.html>, 2003.

アピールチャート



未来ビジョン

爆発的に増加しつつある情報からどのような有益な情報を得るか. このような問題意識を前提とした研究は多い. Google がビジョンとして「世界中の情報を整理し, 誰にでもアクセスできるようにする」を掲げてから数年, インターネットの本格的な普及が始まってから 10 年以上が経過した.

かつての「未来」が部分的にであっても「現実」となった今, 我々は今一度自分たちの姿を振り返り, そうした研究, 発達の方向が本当に我々に幸せをもたらしているのか自問すべきではないだろうか. 大量の情報を保存し知識を増やしたつもりになり, 反射的に思いついた言葉を撒き散らす. これが我々の目指した未来だろうか. もたらされたのは情報という名の阿片を吸った中毒者が見る夢ではないのか.

本論文で使われている個々の技術について

の最適性を主張するつもりはない. また解決策として向かうべき方向性についても最適性を主張しない.

なによりも主張したいのは, 「有益な情報を得られること = 幸せな生活の実現」というシンプルな前提に疑問を持ち, 考え治すことの必要性である.

未来を切り開くのが研究者の役割の一つだとすれば, 自分たちがどのような未来に向かっているのかについて考察し, 責任をもつことも求められるのではないだろうか. ロボット研究における目的の虚構性を指摘した文章 [16] は, インタラクション研究にも等しく当てはまる. 向かっている未来に疑問を持ち, 新たな研究の方向性を持つことで小さくはあっても変化を起こすことができる, と考えている.

本論文がそうした「今我々が向かっている未来」の先行きについて考察, 議論をするきっかけとなれば幸いである.