

推薦論文

スポーツの勝敗にまつわるネタバレ防止手法： 情報曖昧化の可能性

中村 聡史^{1,a)} 小松 孝徳¹

受付日 2012年6月27日, 採録日 2013年1月11日

概要：スポーツの試合の録画視聴を楽しみにしているユーザーにとって、Web上で遭遇してしまう試合結果などのネタバレ情報は楽しみを減退させる忌むべきものである。本稿では、こうしたネタバレ情報との遭遇を防ぐための表現手法を4つ提案し、実装を行った。また、表現手法の有効性について評価実験をベースとして検討を行った。

キーワード：情報曖昧化, ネットバレ, スポーツ, ウェブ

Study of Methods to Diminish Spoilers of Sports Match: Potential of a Novel Concept “Information Clouding”

SATOSHI NAKAMURA^{1,a)} TAKANORI KOMATSU¹

Received: June 27, 2012, Accepted: January 11, 2013

Abstract: Seeing the final score of a sports match on the Web often spoils the pleasure of a user who is waiting to watch a recording of this match on TV. This paper proposes four information clouding methods to block spoiling information, and describes implementation of a system using these methods as a browser extension. We then experimentally investigate the usefulness of the methods, taking into account their differences, differences in the variety of content, and differences in the user's interest in sports.

Keywords: information clouding, spoilers, sports, web

1. はじめに

スポーツは筋書きのないドラマであるため、録画のコンテンツであっても結果や経過を知らなければハラハラドキドキしつつ展開を楽しむことができるが、結果を知ってしまうと展開や結果に対する安心感があるため、楽しみ方が変わってしまう。たとえば、推理ゲームを楽しもうとしているユーザーにとって、そのゲームの犯人の名前やトリックなどを知ってしまうと、淡々とこなすだけになってしまい、驚きや試行錯誤などといった楽しみはなくなってしまう。その一方、映画やドラマなどは、視聴者を楽しませるため

の演出が様々に存在し、視聴者は思いもしない展開を楽しむことができる。しかし、演出や展開が分かってしまうと、面白さは減退してしまうであろう。実際に、アメリカの人気テレビドラマ glee の収録に参加していたエキストラが、そのドラマの放送前にそのサプライズ企画を Twitter で発信してしまい、解雇されるなど大きな問題となった*1。

以上のように、ある対象コンテンツをまだ楽しんだことがなく、これから楽しもうと思っているユーザーにとって、その対象に関する経過や結果などのネタバレ情報は望ましいものではない。もっとも、ネタバレ情報を知っていても楽しめるようなコンテンツは存在しており、歴史ドラマや

¹ 明治大学総合数理学部先端メディアサイエンス学科
Department of Frontier Media Science, School of Interdisciplinary Mathematical Sciences, Meiji University, Nakano, Tokyo 164-8525, Japan

^{a)} satoshi@snakamura.org

本稿の内容は 2012 年 3 月のインタラクション 2012 で報告され、同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

*1 <http://www.tvguide.com/news/glee-prom-spoilers-1032011.aspx>

スポーツの名勝負などがあげられる。また、応援しているチームが勝利した試合だけ視聴したいといったように、安心感をもってコンテンツを楽しみたい人にとっては、ネタバレ情報は問題のないものである。我々は、こうした安心感をもってコンテンツを楽しむようなユーザは対象とせず、ネタバレ情報との出会いをそのコンテンツを楽しむまで極力避け、ハラハラしながら楽しみたいと考えるユーザを対象としたネタバレ防止システムの実現を目指している。なお、本稿ではコンテンツとしてスポーツを対象とする。

スポーツを楽しもうとしているユーザは、基本的にリアルタイムでの視聴を望んでいる。しかし、勉強や仕事、私的な用事、放送局や時差の都合などによりリアルタイム視聴ができないことが多々ある。ここで、Golbeck [10] も言及しているとおり、ネタバレ情報から距離をおこうとしているユーザは、その情報に関する完全な遮断を望んでいる。ネタバレ情報の遮断は、テレビを視聴せず、ウェブを閲覧せず、他人との接点をなくすことで可能である。しかし、日常生活の中で人間は様々な活動をしているため、完全な遮断は難しい。たとえば、情報遮断のため閲覧するウェブページなどを制限していたとしても、重大ニュースについて断片的に情報を得たときに、詳細な情報を得ようとニュースポータルなどにアクセスし、ネタバレ情報に触れてしまうこともある。また、生活の中で遭遇した面白い出来事について Twitter などのコミュニケーションサービス上に発信をしようとして他者の流したネタバレ情報に遭遇してしまうことや、日常的に巡回しているブログなどでメインコンテンツの横に新着記事として試合結果などのネタバレ情報が表示され、意図せずネタバレに遭遇してしまうこともある。最近では、Google などの検索サービスではニュースも検索対象となっているため、ウェブページ検索を行った際に、検索結果の上位に試合結果などが表示され、ネタバレ情報が目に入ってしまうこともある。

そこで我々は、ウェブページに含まれるユーザにとってネタバレとなるテキスト情報を検出し、そのテキスト情報に対する曖昧化処理を行うことによって、ユーザが不意にネタバレ情報と遭遇してしまうことを防ぐ表現手法を提案する。具体的に本稿では4つの表現手法を提案および実装し、その表現手法間の違い、コンテンツ間の違い、またユーザの興味の傾向の違いに注目し、表現手法の有効性を実験的に検証する。

なお、本稿では、システムのネタバレ情報の検出についてはシンプルな正規表現を利用し、精度としての網羅率(再現率)を高めることを主目的とはせず、網羅率が高められない状態であっても、いかにしてユーザにネタバレ情報を悟られないかという点に主眼を置いている。つまり本稿での実験条件は網羅率が一定であると仮定したうえで、表現手法の違いの影響のみを考察するものである。

2. 関連研究

ネタバレ防止に関連した研究として、情報推薦や情報フィルタリングに関する研究があげられよう。膨大な情報の中からユーザの興味や嗜好に基づいて適切な情報や商品などを提示するために、情報推薦 [3], [4] に関する研究が数多く行われている。ここで情報推薦の場合、サッカーの日本代表が好きなユーザに対してはサッカーの日本代表に関する記事や結果を積極的に届けるため、単に情報を手に入れたと思うユーザにとっては有用であるが、情報遮断を行おうとしているユーザにとっては余計なお世話であり、ネタバレ情報に触れる危険性を上昇させてしまう。

情報のフィルタリングに関する研究やシステム開発も多数なされている [5], [6]。こうした研究やシステムでのフィルタリングの対象は、基本的にコンピュータに害をなすウイルスや、日々届く SPAM メール、子どもにとって有害とされる暴力的・性的コンテンツなどのようなものであり、ほとんどのユーザにとって忌むべき不要なものである。しかしこれらの手法は、本研究で対象とするようなユーザの興味に応じた動的なフィルタリングには適していない。

k-Anonymity [7] はデータベースのデータを各種の視点での分析を可能としつつも、個を識別可能な部分について曖昧性を持たせることでプライバシーを保護するための分析および変換手法である。k-Anonymity を利用することで「広島」や「阪神」といった個別のチーム名を「チーム」などの曖昧な情報に置き換え、どのチームが勝利したのかなどを分からなくすることが可能である。ただし、データベース上のデータのような正規化されたデータを対象としており、ウェブ上の雑多なデータへの適用は困難である。

ネタバレ情報が小説を楽しもうとする際にどこまで影響を与えるかということ調査したものに Leavitt らの研究がある [8]。この研究では、ネタバレ情報を提示したときと提示していないときでユーザのコンテンツの楽しみ方どのような差があるのかを実験的に調査しており、その結果からネタバレ情報は実際のコンテンツの面白さを低減させないと主張している。しかしこの実験は、あらかじめあらすじを知っておくことで、小説を読みやすくなるということを明らかにしているにすぎない。これは、本提案で問題としているネタバレとは本質的に異なる。

Ikeda らは機械学習の手法を用いて映画のレビュー記事にあらすじが含まれているかどうかを判定し、そのレビュー記事からあらすじを段階的に除去する研究に取り組んでいる [9]。あらすじが直接ネタバレとなるわけではないため、本稿で取り組んでいる内容とは異なる。

Golbeck は Twitter 上で流れてくるネタバレ情報を防ぐためのネタバレ情報判定手法を提案している [10]。ここでは、適合率(ネタバレ情報として検出したものの正解の割合)ではなく再現率(ネタバレ情報の網羅度合い)を重視

し、ネタバレ情報の判定を行う手法を提案している。再現率を重視するという点で我々のコンセプトにとっても近いものであるが、Golbeckが判定手法に注目して再現率100%を目指しているのに対し、我々は再現率が100%に至らなくても表現手法によってネタバレとの遭遇を避けることができるという点で異なる。

3. ネットバレ情報と従来の遮断手法

一般的に、ポータルサイト、検索サービス、コミュニケーションサービスなどにおいて掲載される速報記事はネタバレ情報に相当するものの、基本的には良かれと思って提示されているものであり、情報遮断を行っていないユーザにとっては価値の高いものであることが多い。一方、そうした情報は、その情報遮断を行っていたユーザが、情報遮断を行っていないときには積極的に集めたい情報であることも多い。たとえば、ワールドカップで日本代表を応援しており、その録画映像を楽しむまで情報遮断をしていたユーザは、そのコンテンツを視聴して日本代表が勝利したことを楽しむと、その楽しみおよび喜びを増幅させるため、ニュースサイトなどを巡回して積極的にそうした情報を収集するであろう。つまり、常時コンテンツを遮断することが望まれているわけではなく、一時的な情報遮断が必要とされているといえる。

我々は過去の研究において、ユーザの興味に基づくネタバレ情報の動的なフィルタリングを可能とする手法を提案し、ローカルプロキシとして実装した [1]。この研究では、ユーザが情報を遮断する対象と設定した遮断対象コンテンツが開始されてから、そのコンテンツ自体をユーザが楽しみ終わるまで、ネタバレ情報をフィルタリングするものである。ここで、筋書きのないスポーツの試合において、試合前に結果を知っているということはありえない。そこで、スポーツの試合の場合は、試合が開始されてから、録画した試合を視聴し終わるまでがフィルタリング期間となり、遮断対象コンテンツ開始後に作成されたコンテンツのみをフィルタリング処理の対象とした。実際のフィルタリング処理では、ウェブコンテンツをブロックに分割し、ブロックがネタバレしているかどうかを正規表現辞書を利用して判定し、そのブロック部分のテキストと背景色を同じ色に設定することで、ネタバレ情報のフィルタリングをしていた。なお、遮断対象コンテンツの登録は、ToDoへの登録またはメールの自動判定により行っていた。

しかし、このシステムはプロキシとしての実装であったため動作が遅いという問題があった。さらに、フィルタリングのためのターゲットコンテンツ登録を重視しているだけであり、ユーザに対してどのように情報提示を行うかという点には考慮していなかった。本研究ではこの手法を発展させ、ブラウザベース上でネタバレを曖昧な情報に変換することによって、ユーザとネタバレ情報との遭遇を防ぐ

ことを目的とし、同時に「情報曖昧化」というコンセプトを提案するものである。ここで情報曖昧化とは、オリジナルの情報をそれと分からないように曖昧なものへと変換することを意味する。

4. 提案手法およびシステム

4.1 ネットバレ対象の登録

家電の情報化が進み、ホームネットワークで相互接続および情報の共有が可能になると、録画したコンテンツの中で何を視聴していないかをシステムが判断し、自動的にネタバレ防止対象を切り替えることができるようになるであろう。現在、そうした環境が世の中に普及しているわけではないため、本稿ではネタバレ防止対象の登録方法については、ユーザにそのネタバレ防止対象試合の開始時間および、その試合に関連するキーワードをオブジェクト名リストとしてユーザに入力してもらう。

ここで、オブジェクト名リストは、チーム名および選手名（正式名称やニックネームなど）からなる。オブジェクト名リストを毎回入力するのは手間であり、負荷がかかってしまう。そこで、ユーザがその試合に該当するチームや選手、試合に関する各種の記述がなされたウェブページ（Wikipediaのページなど）を複数指定すると、システムがMeCabを利用して形態素解析を行い、自動的にその中から固有名詞と未知語を抽出し、その語を利用可能とする方法を用意している。ユーザは抽出されたオブジェクト名リストのチェックを外すことにより、該当しないキーワードを排除することが可能である。そのため、指定するウェブページがつねに最新のものに更新されていれば新しい選手などへの対応は可能である。なお、今後は様々なスポーツのチームや選手を管理するデータベースおよびアクセスAPIを用意し、ネタバレ防止のための試合情報を共有するサービスを実現予定である。

4.2 ネットバレ情報の検出

ユーザが閲覧しようとしているコンテンツ自体が情報遮断対象の試合の開始以前に最終更新されている場合、システムはそもそも処理を行わない。コンテンツの最終更新日が不明の場合（CGIなど）や、コンテンツが情報遮断対象の試合の開始以降に更新されている場合、そのコンテンツにネタバレ情報が含まれているかどうかを検証する。テキスト部分がハイパーリンクとなっている場合、そのリンク先のコンテンツの最終更新日時を確認して、その部分のネタバレ度合いを検証する。

ネタバレ情報が含まれているかどうかの検証の前段階として、まずコンテンツをブロック単位で認識し、そのブロックがネタバレ情報を含んでいるかどうかで判定する。また、そのブロックから他のコンテンツへとリンクが設定されている場合には、そのリンク先のコンテンツの更新日

/[オブジェクト][助詞]*[0-9]+[*]-:対][*][0-9]+/
/[オブジェクト][0-9]+[*]-:対][*][0-9]+[助詞]*[オブジェクト]/
/[オブジェクト][助詞]*[勝利]圧勝[優勝]敗北[敗退].../
/[勝利][優勝]敗北[敗退]...[助詞]*[オブジェクト][助詞]*/
/[オブジェクト][助詞]*V[0-9]+/
/[オブジェクト][助詞]*[0-9]+[連勝]連敗.../

図 1 正規表現の一例 (書き言葉用)
Fig. 1 Examples of regular expression.

時を取得し、その更新日時が対象イベント開始日時より古いものであればネタバレ情報でないと判定し、新しいものである場合はコンテンツ自体を取得し、そのコンテンツのネタバレ可能性について検証する。なお、ブロックの認識については、視覚的な特性を利用するもの [2] など様々な手法が提案されているが、本研究では単純のため、タグで囲まれた範囲を 1 つのブロックとする。

ブロック単位でのネタバレの有無の検証においては、情報遮断対象を識別するためのオブジェクト名リストと、ネタバレ情報を検出するための勝敗にまつわる正規表現辞書を利用する。正規表現辞書の例は図 1 のとおりである。また、結果を変換するための、結果変換辞書も用意している (圧勝の変換語は辛勝, 惜敗, 完敗. ゴールの変換語はミスやオウンゴール, 優勝の変換語は準優勝, ベスト 4 など). なお, 現時点では, 勝敗の判定に関しては 73 の正規表現を, 結果の反転に関しては 154 の正規表現を使用している。スポーツについてはサッカーだけでなく, バスケットボールやバレーボール, 野球や自転車競技などメジャースポーツはカバーしているが, 体操や F1 などについては現時点では十分にカバーされていない。

4.3 情報曖昧化手法

本研究では, 下記の 4 つの情報曖昧化のための表現手法を提案する。

- **非表示手法**: ネタバレが含まれる該当ブロック部分を単純に非表示に設定する表現手法。ただ単にその部分が抜け落ちるだけであるため, ユーザはそのページを閲覧中に気になることはないが, その対象が遮断すべきものでなかった場合であっても, ユーザはそのコンテンツを調べることができない。
- **墨塗り手法**: ネタバレが含まれる該当部分の背景およびテキストを同じ色へ設定することで, 遮断対象部分が, 検閲により墨塗りされたように視覚化される表現手法。どこが曖昧化されているかをユーザは把握可能である。遮断されたテキストもマウスで選択することにより閲覧可能である一方で, ユーザはその部分が気になってしまうと予想される。
- **木の葉を隠すなら森の中手法**: 結果変換辞書を利用して, ネタバレ情報と結果が反転したものや, 類似した結果などをページ中に挿入していく表現手法。本手法

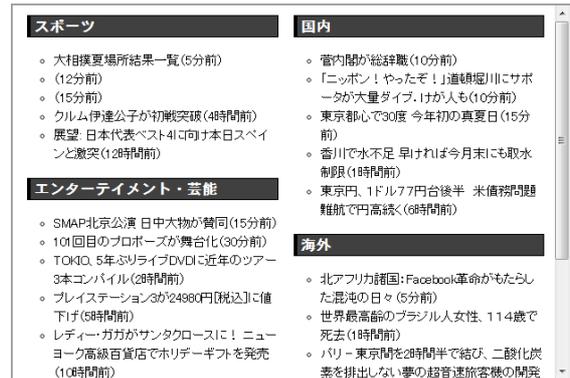


図 2 非表示手法
Fig. 2 Invisible method.

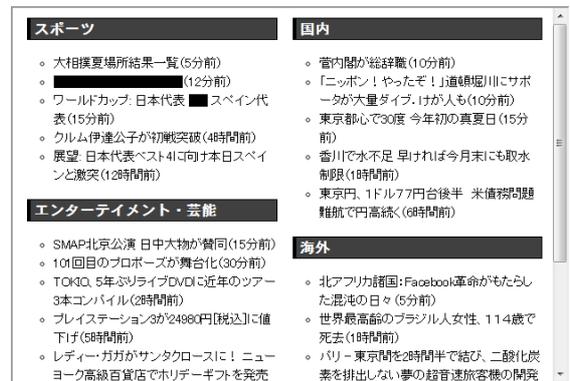


図 3 墨塗り手法
Fig. 3 Censored method.

により, ユーザはどの情報が正しいオリジナルのものが分からなくなる。ネタバレとの遭遇の可能性は低くなるが, ゴミだらけになってしまい鬱陶しく感じる可能性がある。以下, 木の葉手法と表記する。

- **結果反転手法**: ネタバレ情報が明確に記述されている部分を, 結果変換辞書を利用して, 勝敗を反転させた記述に変更する表現手法 (「日本代表が勝利」を「日本代表が敗北」「日本代表が惜敗」などに置換)。ユーザは遭遇したネタバレ情報がそもそも正しいものかどうかを判断できない。ただし, ネタバレ防止対象はユーザが登録するため, 情報を見た後でその情報は曖昧化されているのではと判断することができる。システムのネタバレ検知精度が低くても (再現率が低くても) 本手法は有効に働く可能性がある。

4.4 実装

提案手法に基づき, Mozilla Firefox の拡張として JavaScript および XUL を用いてシステムを実装した。図 2 は, 非表示手法を用いて情報の曖昧化を行っている様子であり, 図 3 は同じくサッカーワールドカップに関する試合について墨塗り手法で情報曖昧化を行っている様子である。また, 図 4 は木の葉を隠すなら森の中手法で情報曖

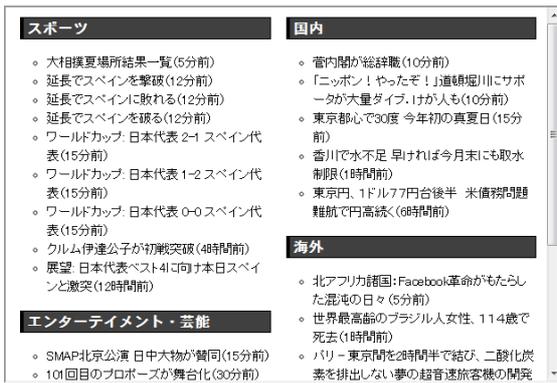


図 4 木の葉手法

Fig. 4 Increasing method.



図 5 結果反転手法

Fig. 5 Inverse method.

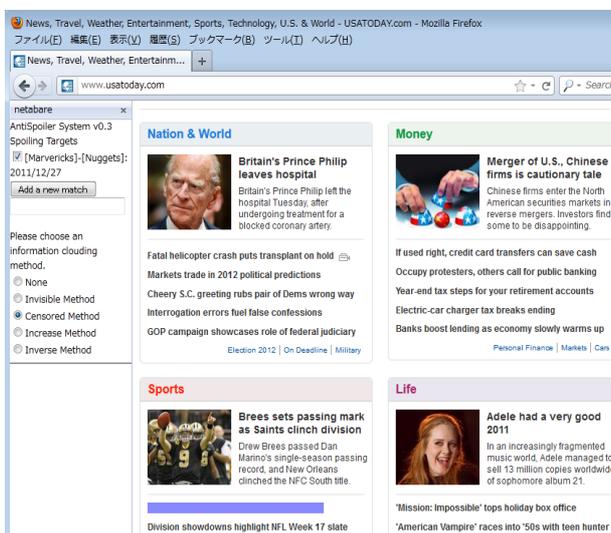


図 6 システムのスクリーンショット

Fig. 6 A screen shot of our system.

曖昧化を行っている様子、図 5 は結果反転手法によって情報曖昧化を行っている様子である。

図 6 はシステムのスクリーンショットである。ブラウザウインドウの左側にネタバレ防止システムの操作 UI が表示されている。曖昧化を施す対象コンテンツの指定については、この UI から入力を行う。また、イベントを楽しん

だ後には、このインタフェースで試合からチェックを外すことにより、そのイベントの終了を伝達する。なお、ユーザは情報曖昧化の表現手法の種類をこの操作システムで切り替えることも可能である。

5. 評価実験

5.1 実験概要

前章で提案した 4 つの表現手法の有効性を検討する実験を実施した。具体的には、表現手法の違い、コンテンツの違い、ユーザのスポーツに対する興味度合いの違いという 3 つの要因が、ネタバレ防止システムに対するユーザの評価にどのような影響を与えているのかを検討することが目的である。

5.2 参加者

本実験には 58 人 (男性: 40 人, 女性: 18 人, 21~27 歳) の大学生および大学院生が参加した。これらの参加者は、インタラクションおよび Web 関係の研究を実施している大学の研究室に所属しており、それらの研究室に向けた実験参加者募集依頼に応じることで本実験に参加した。なお、実験参加に対する報酬は用意されなかった。

5.3 実験手順

本実験は、我々によって開発されたオンライン実験システムで実施された。実験システムは、図 2~図 5 のように 4 つの表現手法を適用した 4 つのニュースポータルページのコンテンツが提示されるというものである。コンテンツは架空のスポーツの結果および架空の重大ニュースを含んでおり、スポーツの結果としては、「延長でスペインを撃破」のように直接的に結果が分かり、システムがネタバレであると検知しやすいものと、「ニッポン! やったぞ!」のように間接的に結果が分かるがシステムがネタバレと検知しにくいものの 2 種類が用意された。本実験におけるスポーツコンテンツとしては、ワールドカップのベスト 4 をかけた男子サッカー日本代表、ロンドン五輪の予選突破を目指す女子サッカー日本代表、アメリカとの決勝に臨む WBC 野球日本代表、日本人初のウィンブルドン制覇を目指すテニスの錦織選手、という 4 種類が用意された。具体的な実験の手順を下記に示す。

- 参加者が実験システムの URL にアクセスすると、まず実験に関する実験許諾書が画面に表示される。それを許諾すると、「今日は非常に重要なスポーツイベントが開催されるが、あなたはそれをリアルタイムには見ることができないため、その試合を録画して帰宅後に楽しむことを考えている。帰宅までその試合の情報を遮断しながらも通常業務を必要があるので、自分の使用しているウェブブラウザにネタバレ防止システムを導入している」という架空の状況説明がなされる。

表 1 システムに対する質問項目：質問 5, 7 は逆転項目

Table 1 Questionnaire for each method of information clouding.

1	ネタバレ防止ブラウザが役にたっている
2	知りたくない情報が隠されているので安心だ
3	どの情報が正しくてどの情報が正しくないのかわからなくなってしまう
4	隠べいされた個所が逆に気になって仕方がない
5	知りたくない情報がそれとなく分かってしまう
6	この手法を早く実現してほしい
7	どのような情報が隠べいされているのかわからない

- その状況下において、国内外の重大な架空のニュースに関する情報が偶然耳に入ったので、ニュースサイトにアクセスして、なるべくネタバレに関する情報を避けつつ、重大なニュースとは何かを把握してほしいとのタスクが与えられる。参加者にはネタバレ防止のための表現手法のうち、いずれかの表現手法が施されたニュースサイトのポータルページを閲覧してもらい、重大ニュースについて報告してもらおう。その際、スポーツの試合結果が分かってしまった場合はその結果は何だったのかも記述してもらおう。そして、ネタバレ防止システムの評価に関する7問(表1)について、7段階のリッカート尺度(最低点1点, 最高点7点)で回答してもらおう。
- この手順を各表現手法4種類について繰り返した後、参加者のスポーツに対する興味度合いを7段階のリッカート尺度(最低点1点, 最高点7点)で採取して実験を終了する。

すべての参加者は4つの表現手法すべてを経験するが、その順序はランダムになるように配慮され、またその際に、同一のスポーツコンテンツが現れないように配慮した。

5.4 評価方法

本実験の結果は、表現手法間要因(4水準;非表示/墨塗り/木の葉/結果反転)、コンテンツ要因(4水準;男子サッカー/女子サッカー/野球/テニス)、スポーツ興味要因(被験者間配置, 2水準;スポーツに対する興味度合いが高い/低い)という3つの独立変数について解析された。なお、スポーツ興味要因は、実験終了間際に採取した「あなたはスポーツに興味がありますか」という7段階リッカート尺度の質問項目において1~3点という低めの得点をつけた27人を「スポーツに対する興味度合いが低い群(以降, 興味低群)」, 4~7点という高めの得点をつけた31人を「スポーツに対する興味度合いが高い群(以降, 興味高群)」と分類した。そして、システムに対する質問項目7問それぞれの得点および参加者が「試合結果が分かった」と報告し、その報告が正解していた数(ネタバレ情報に気づいた数。以下, 正解数)の2つを従属変数として扱うこととした。

本実験では、表現手法要因およびコンテンツ要因を組み

合わせた16種類の実験刺激が用意されているものの、各参加者はこれらのうち4種類のみをランダムな順番で経験するために、この2つの要因をまとめて分析を行うことはできない。そこで、これら2つの要因はスポーツ興味要因と組み合わせ分析することとした。具体的には、「表現手法要因およびスポーツ興味要因が従属変数に与える影響」「コンテンツ要因およびスポーツ興味要因が従属変数に与える影響」の2点に着目した分析を行った。

5.5 結果

5.5.1 表現手法要因およびスポーツ興味要因の影響

まず各質問項目に関して、二要因混合計画(独立変数その1:表現手法要因(被験者内配置, 4水準), 独立変数その2:スポーツ興味要因(被験者間配置, 2水準), 従属変数:質問項目における得点)の分散分析を行った。分析結果を図7および表2に示す。これらの結果より、全7問の質問項目すべてにおいて表現手法要因に有意差が観察された。よって、本稿で提案した4種類の表現手法は、参加者にとって様々に解釈されていることが明らかになったが、全般的には、木の葉および結果反転手法よりも非表示および墨塗り手法の方が高い評価を受けていると示唆された。また質問1, 2において、表現手法要因および興味要因の交互作用に有意差および有意傾向が観察された。つまり、この2項目において、スポーツに対する興味の高低が各表現手法に対する評価の差異につながっていることが明らかになった。具体的には、興味高群の参加者は非表示手法を好んでおり、興味低群の参加者は墨塗り手法を好んでいることが明らかになった。質問6において、スポーツ興味要因に有意差が観察されたこともあわせて確認された。つまり興味高群の参加者は、興味度合いが低い参加者に比べ、このようなネタバレ防止システムを欲していることが明らかになった。

続いて、コンテンツの内容がこれらの結果にどのような影響を及ぼしているかを観察するために、コンテンツごとに表現手法要因とスポーツ興味要因の二要因の影響を分析した。各質問項目に対して、二要因被験者間計画(独立変数その1:表現手法要因(被験者間配置, 4水準), 独立変数その2:スポーツ興味要因(被験者間配置, 2水準), 従属変数:質問項目における得点)の分散分析を行ったところ、コンテンツごとの解析には表現手法ごとの解析と比べて下記のような特徴的な結果が観察された。

- サッカー男子:質問1, 2には交互作用は観察されなかった(質問1: $F(3, 57) = 0.97$, n.s., 質問2: $F(3, 57) = 1.24$, n.s.).
- サッカー女子:全体的な傾向と類似した結果が観察された。
- 野球:質問3のスポーツ興味要因に有意差が観察された(興味高群 > 興味低群: $F(1, 57) = 5.04$, $p < .05$).

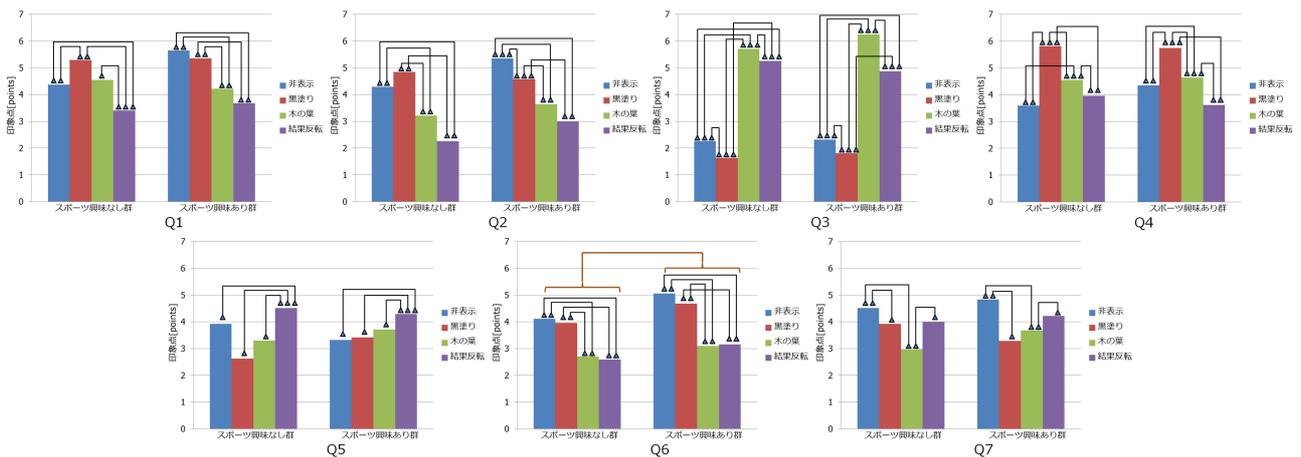


図 7 表現手法要因およびスポーツ興味要因の二要因混合計画の分散分析結果：項目間の実線は 5%水準での有意差を示している

Fig. 7 Results of mixed ANOVA among method and preference factors.

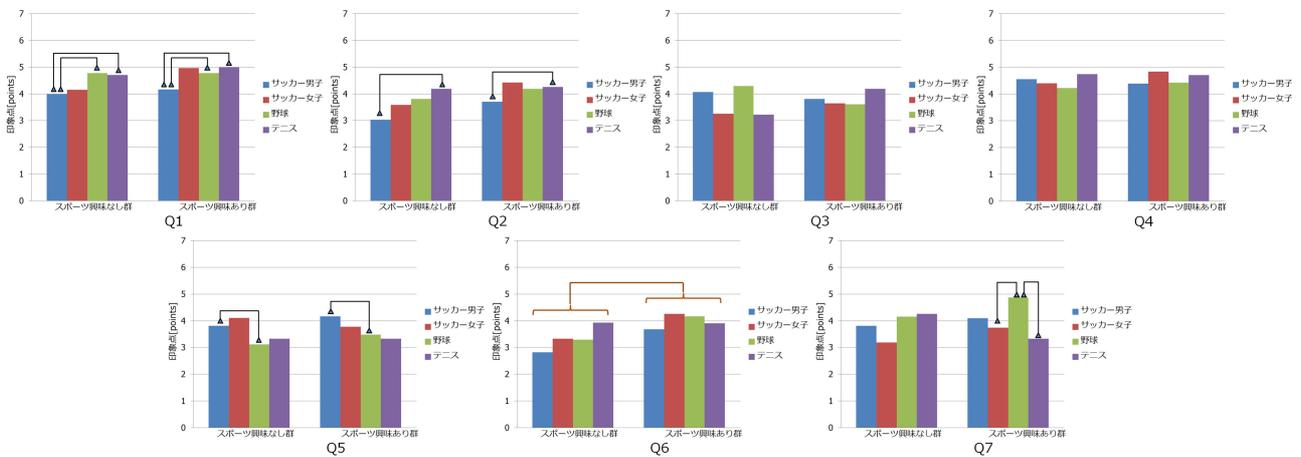


図 8 コンテンツ要因およびスポーツ興味要因二要因混合計画の分散分析結果

Fig. 8 Results of mixed ANOVA among contents and preference factors.

表 2 表現手法要因・スポーツ興味要因の二要因混合計画の統計量

Table 2 F-statistical of mixed ANOVA among method and preference factors.

	スポーツ興味要因	手法要因	交互作用
Q1	F (1,56) = 1.68, n.s.	F (3,168) = 15.25, p < .01	F (3,168) = 2.89, p < .05
Q2	F (1,56) = 2.84, p < .10	F (3,168) = 32.84, p < .01	F (3,168) = 2.37, p < .10
Q3	F (1,56) = 0.22, n.s.	F (3,168) = 119.73, p < .01	F (3,168) = 1.04, n.s.
Q4	F (1,56) = 0.10, n.s.	F (3,168) = 17.29, p < .01	F (3,168) = 1.20, n.s.
Q5	F (1,56) = 0.11, n.s.	F (3,168) = 5.96, p < .01	F (3,168) = 1.78, n.s.
Q6	F (1,56) = 5.60, p < .05	F (3,168) = 28.94, p < .01	F (3,168) = 0.44, n.s.
Q7	F (1,56) = 0.31, n.s.	F (3,168) = 6.30, p < .05	F (3,168) = 1.43, n.s.

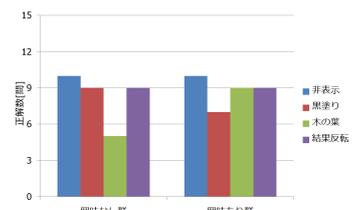


図 9 手法要因ごとの正解数

Fig. 9 Spoiled numbers in each method.

- テニス：質問 3 および 7 の表現手法要因のみに有意差が観察された (質問 3 : $F(3,57) = 31.31, p < .01$, 質問 7 : $F(3,57) = 4.19, p < .05$).

これらの結果より、サッカー男子は他のコンテンツと比べて質問 1 および 2 の非表示、黒塗り手法に対する回答傾向が異なっていることが明らかになった。また、他のコンテンツと比較すると、テニスに関しては表現手法要因の影響が小さかったことが明らかとなった。

参加者が実際にネタバレに気づいた数を表現手法要因ごとにまとめた「正解数」を図 9 に示す。スポーツ興味要

因間で同じ表現手法のペアを比較およびスポーツ興味要因内でそれぞれの表現手法のペアを比較する直接確率計算を行ったところ、いずれのペアについても有意差が観察されなかったため、表現手法要因およびスポーツ興味要因はユーザがネタバレ情報を理解したか否かという機能的な側面に関しては特に影響を与えていなかったことが確認された。

5.5.2 コンテンツ要因およびスポーツ興味要因の影響

各質問項目に関して、二要因混合計画 (独立変数その 1 : コンテンツ要因 (被験者内配置, 4 水準), 独立変数その 2 :

表 3 コンテンツ要因・スポーツ興味要因の二要因混合計画の統計量
Table 3 F-statistics of mixed ANOVA among contents and preference factors.

	スポーツ興味要因	コンテンツ要因	交互作用
Q1	F(1,56) = 1.68, n.s.	F(3,168) = 2.36, p < .10	F(3,168) = 0.62, n.s.
Q2	F(1,56) = 2.84, p < .10	F(3,168) = 2.55, p < .10	F(3,168) = 0.52, n.s.
Q3	F(1,56) = 0.22, n.s.	F(3,168) = 0.50, n.s.	F(3,168) = 1.19, n.s.
Q4	F(1,56) = 0.10, n.s.	F(3,168) = 0.51, n.s.	F(3,168) = 0.28, n.s.
Q5	F(1,56) = 0.11, n.s.	F(3,168) = 2.40, p < .10	F(3,168) = 0.47, n.s.
Q6	F(1,56) = 5.60, p < .05	F(3,168) = 2.07, n.s.	F(3,168) = 1.24, n.s.
Q7	F(1,56) = 0.31, n.s.	F(3,168) = 3.28, p < .05	F(3,168) = 2.42, p < .10

スポーツ興味要因（被験者間配置，2水準），従属変数：質問項目における得点）の分散分析を行った結果を図8および表3に示す。これらの結果より，表現手法要因の影響と同様に，質問6においてスポーツ興味要因に有意差が観察され，興味高群は，興味低群に比べてこのようなシステムを欲していることが改めて明らかになった。そして質問1, 2, 5, 7で，コンテンツ要因に有意傾向が観察され，特にサッカー男子に関するコンテンツに対しての評価が低い傾向にあるということが示唆された。さらに質問7には交互作用にも有意傾向が観察され，テニスというコンテンツに対して興味高群の評価の方が低く，さらに各コンテンツに対しての評価にも差異が生じていたことが明らかになり，具体的には，野球に対する得点が，サッカー女子およびテニスよりも高かったことが明らかになった。

続いて，表現手法の違いがこれらの結果に何らかの影響を及ぼしているか否かを観察するために，表現手法ごとにコンテンツ要因とスポーツ興味要因の二要因の影響を分析した。各質問項目に対して，二要因被験者間計画（独立変数その1：コンテンツ要因（被験者間配置，4水準），独立変数その2：スポーツ興味要因（被験者間配置，2水準），従属変数：質問項目における得点）の分散分析を行ったところ，表現手法ごとに以下のような特徴的な結果が観察された。

- 非表示手法：質問1, 2, 6で，スポーツ興味要因に有意差が観察された（興味高 > 興味低：質問1：F(1,57) = 7.21, p < .05, 質問2：F(1,57) = 4.20, p < .05, 質問6：F(1,57) = 6.22, p < .05）。
- 墨塗り手法：質問1, 5で，コンテンツ要因に有意差が観察された（質問1：F(3,57) = 3.15, p < .05, 質問5：F(3,57) = 4.63, p < .05）。
- 木の葉手法：質問1に交互作用に有意差が観察された（F(3,57) = 3.29, p < .05）。
- 結果反転手法：質問3にのみコンテンツ要因に有意差が観察された（F(3,57) = 3.06, p < .05）。

この結果より，非表示や墨塗りといった表現手法はスポーツに対する興味やコンテンツの違いがその評価に影響を与えていることが見受けられる一方，木の葉や結果反転といった表現手法はそのようなムラが存在していないことが示唆された。

また，コンテンツ要因ごとの正解数を図10に示す。こ

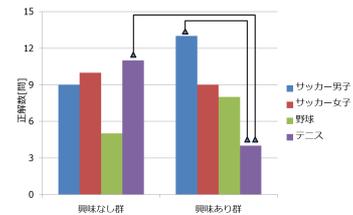


図 10 コンテンツ要因ごとの正解数

Fig. 10 Spoiled numbers in each content.

の図より，スポーツ興味高群においては，コンテンツごとの正解数にばらつきが大きいことが見受けられた。そこで，スポーツ興味要因内でそれぞれのコンテンツのペアを比較する直接確率計算を行ったところ，興味高群内で，サッカー男子とテニスとの間，またサッカー女子とテニスとの間に有意差が観察された。また，スポーツ興味要因間で同一のコンテンツを比較する直接確率計算を行ったところ，テニスにおいて有意差があることが観察された。

6. 考察

6.1 提案した情報曖昧化手法に関する考察

本実験の結果，本稿で提案した4つの曖昧化手法に関しては，非表示手法や墨塗り手法の方が，木の葉手法や結果反転手法に比べ，Q1, Q2, Q6において高い評価を受けているため，非表示手法や墨塗り手法の方が木の葉手法や結果反転手法よりも有効だと考えられていたことが分かった。非表示手法や墨塗り手法が良いと評価された理由の1つは，木の葉手法や結果反転手法だとどの情報が正しくてどの情報が正しくないかということが分からなくなるというものである。この情報の正しさに対する不安感が結果としてこれらの表現手法の悪印象につながっているのではないかと考えられる。つまり，プロ野球のようにほぼ毎日試合があるようなスポーツに対して木の葉手法や結果反転手法を適応すると，日常的に閲覧する情報に対してもネタバレ処理が施されるために，それ以外の記事に対しての信憑性をも下げてしまうという問題が生じてしまっていると考えられる。一方，ワールドカップのサッカー日本代表の試合のように，ここだけはなんとかネタバレを防止したいと思うような場合，木の葉手法や結果反転手法は効果的に働く可能性がある。このような試合の珍しさと重要度を考慮した情報曖昧化手法の比較検討については，今後実験的に検証していく予定である。

今回提案した4つの表現手法のうち，非表示手法および墨塗り手法については，その表示自体が気になってしまううえ，何がフィルタリングされているか分からないという不安感が参加者から報告されていた。特に，墨塗り手法はマウスで該当部分を選択することにより結果を知ることができるが，非表示手法はそれさえも知る手段がないため問題となっていたようである。

6.2 スポーツへの興味度合いと情報曖昧化手法との関係

実験の結果より、興味低群の参加者に比べ、興味高群の参加者は情報曖昧化手法を欲していることが明らかになった。つまり、スポーツにより興味のある参加者は、ネタバレを日常における重要な問題としてとらえており、本研究の需要を示しているといえる。また、興味低群の参加者の方が、サッカー男子というコンテンツに比べてテニスの方が知りたくない情報が隠されているので安心であると回答していたことが観察された。これは、サッカー日本代表の試合はワールドカップなどで認知度を上げており、スポーツにあまり興味がなくても、どういったものがネタバレであるかなどを参加者が理解しているためであると考えられる。つまり、スポーツにあまり興味がなくても、ここぞというときの試合についてはネタバレ防止を求められる可能性を示唆している。

さらに、興味高群の参加者は、興味低群の参加者に比べ、テニスというコンテンツに対してどのような情報が隠蔽されているのか分かってしまうという評価をしていた。しかしその一方、実際のネタバレ数を示す正解数においては興味低群の参加者の方が、興味高群の参加者よりも多いことが確認された。これは、情報曖昧化手法によって、スポーツに高い興味を持つ参加者の「ネタバレである」という予想を外しており、結果としてネタバレを防ぐことができているといえる。

6.3 運用からの知見

提案および実装したシステムを南アフリカワールドカップ期間中に、ワールドカップの各試合を対象として著者および実験協力者1人で運用を行った。コンテンツの曖昧化速度はいずれの表現手法であってもコンテンツ提示から曖昧化までが瞬時に行われるため、ストレスは感じなかったことと報告された。

対象ブロックコンテンツがネタバレ対象であるかどうかについての判定について、今回の判定手法は辞書に登録されているキーワードがブロックに含まれていれば曖昧化対象とするという簡易的なものであった。そのため、全体のネタバレ情報のうち、どれだけ多くのネタバレ情報を曖昧化できたかという、曖昧化判定の再現率は高く、Yahoo!スポーツとmixiニュース、朝日新聞において、用意した108のネタバレ情報のうち、102のネタバレ情報を検出できていた（ブロックをタグ単位にしているため、テーブルタグなどで結果が表示されている場合に検出できないなどの問題もあった）。一方、108のネタバレ情報しかないのに、142のブロックをネタバレ情報として検出してしまうなど、適合率の面で問題があった。これは、今回の判定手法が再現率を重視したものであることが原因の1つである。

今回実現した4つの表現手法のうち、非表示手法および墨塗り手法については、先述の実験結果と同様、その表示

自体が気になってしまううえ、何が曖昧化されているか分からないという不安感があった。一方、木の葉を隠すなら森の中手法は、類似のタグを挿入する手法であるため、サイトによってはページのレイアウトを壊してしまうという問題があった。しかし、うまく変換できた場合にはネタバレ情報の存在を示すとともに嘘の情報も提示されるため、その情報を見た際に対象イベントへの興味が増幅されることがあった。結果反転手法は、結果変換辞書の精度にかかわらず、実際にネタバレしているのかどうか分からないという利点があった。結果が分かってしまったと驚いてしまうこともあったが、ユーザはその情報を曖昧化の対象としていることが分かっているため、実際にはネタバレリスクが低減される。たとえば、「ニッポンやったぞ。道頓堀川にサポーターがダイブ」などのニュースのように、ネタバレ情報の記述がシステムにとって検知しにくいものであった場合、システムがネタバレ情報をそのまま提示してしまう。ここで、非表示手法や墨塗り手法、木の葉を隠すなら森の中手法であれば、その情報が表示されているという時点でユーザはネタバレであることに気づくが、結果反転手法の場合はその情報も変換されたものではないかとユーザが考えるため、結果としてネタバレにならないということがあった。

先述のとおり、スポーツについてはサッカーだけでなく、バスケットボールやバレーボール、野球や自転車競技などメジャースポーツはカバーしているが、体操やF1などについては「○×がクラッシュ」や「○×が落下」などのような結果情報はカバーできていない。現時点では十分にカバーされていない。広くスポーツをカバーすることは難しく、新語も登場する可能性があるため、今後はWeData^{*2}などの仕組みを利用してWeb上で正規表現を共有し、Webブラウザの起動時に正規表現を読み込み利用するような仕組みを実現する予定である。また、Ikedaら[9]のように機械学習の手法を利用することによってメンテナンスフリーの辞書を構築することも検討している。なお、スポーツであれば勝敗に関係するので問題ないが、スポーツ以外のネタバレ（小説や映画、ゲームなど）を対象とする場合は、辞書は分野ごとに用意する必要があると考えられる。こうした点については、今後検討を行う予定である。

新聞社などの報道機関が発信する結果情報に比べ、Twitterや掲示板サービスなどにおいて一般ユーザから発信される結果情報に対する曖昧化に失敗していることが多かった。これは、報道機関が発信する情報の場合、コンテンツは書き言葉で記述されていることが多いが、Twitterなどのサービスでは話し言葉で記述されることが多いためである。また、Twitterなどのように膨大なユーザが高頻度で気軽に発信するようなサービスの場合、単に「ゴール」

^{*2} <http://wedata.net/>



小松 孝徳

1974年生．2003年東京大学大学院総合文化研究科博士課程修了．同年公立はこだて未来大学システム情報科学部助手，2007年信州大学ファイバーナノテク国際若手研究者育成拠点助教，2012年信州大学繊維学部准教授，2013

年明治大学総合数理学部准教授，現在に至る．人間の直感的認知特性という切り口から，人間と人工物との間のインタラクションをとらえる研究活動に従事．日本認知科学会，人工知能学会，ヒューマンインタフェース学会，ACM等の各会員．博士（学術）．