

Web 上でのインフォーマル・コミュニティ範囲 選択方式の検討と評価

山口亮輔[†] 森口俊幸[†] 金井敦[†]

近年,現実世界においてインフォーマル・コミュニケーションと呼ばれる,計画的でなく,カテゴリや話題性が存在しない,偶発的なコミュニケーションが注目を集めている.現在 Web 世界においてもインフォーマル・コミュニケーションを実現するためにいくつかのサービスが提案されている.しかし,既存のサービスでは1つの Web ページに集まっているユーザ間を想定しており,意味的な範囲の広さが考慮されていない.そこで,本稿ではコミュニケーションを行う範囲をインフォーマル・コミュニティ範囲と呼び,その範囲を定義し,その実現方式として URL のディレクトリ階層を利用した手法を検討し,評価を行い,その適応領域を明らかにした.

The study and evaluation on a methodology to select a Informal Community range on the Web

Ryosuke Yamaguchi[†] Toshiyuki Moriguchi[†]
and Atsushi Kanai[†]

Recently, the importance of unscheduled, unarranged, accidental communication (Informal communication) has been recognized in real world. On the other hand, some service is suggested to realize informal communication on the Web. These existing services provide communication range based on only one page. However, it is necessary to communicate in more semantically wide range. In this paper, the range of web pages available to communicate is investigated, and range (We call the range Informal Community range) is defined. The method to grasp the structure of homepages using URL is investigated. The method is evaluated and the area of application is clarified.

1. はじめに

現在, Web 上での情報の共有を目的としたサービスが急激な成長を遂げたことにより,Web を利用した様々なコミュニケーション形態が普及している.

一方,現実世界において,インフォーマル・コミュニケーション[1]と呼ばれる,計画的でなく,カテゴリや話題が存在しない偶発的なコミュニケーションが注目を集めている.これは偶然の出会いから生まれるコミュニケーションであり,現状の Web 世界ではこのインフォーマル・コミュニケーションは起こりえない.その理由として,現在の Web では誰かとコミュニケーションしたいときに,自分の意思でコミュニケーションできる機能をもった Web ページを訪れなければ,人と人が偶然に出会うことも,互いを認識することもできないからと考えられる.そこで,Web 上でインフォーマル・コミュニケーションを実現する方式として,今現在同一ページを閲覧している他者をインフォーマル・コミュニティとし,コミュニケーションすることができる wegru[2]や SPEEKY [3],Nesutu[4]が提案されている[5][6].

現在の wegru や SPEEKY,Nesutu では1つの Web ページに集まっているユーザを想定しているため,ホームページの構成や内容によって異なる適切なコミュニケーションの範囲が考慮されておらず,インフォーマル・コミュニティの範囲に限られる.そこで,ある Web ページを中心とした Web ページグループを構成し,グループ内の他者を認識して,コミュニケーションを行う機能を実現するにあたり,その範囲を制限するために Web ページ間の距離を規定する必要がある.そのために,ホームページの構造を把握する実現方式を検討し,評価を行い,その適応領域を明らかにした.

2. コミュニケーション範囲の規定

wegru には,現在閲覧している Web ページ上でユーザ同士が互いを認識でき,コミュニケーションできる機能が提案されている.しかし,図1に示すように,Web ページによっては,他の Web ページも考慮することで幅広いインフォーマル・コミュニティが形成できると考える.そこで Web ページの適切なコミュニケーション範囲を想定する必要がある.

Web ページの範囲を考える上で現実世界の様子を考察すると,現実世界におけるコミュニケーションの範囲は人との距離や声量といった,物理的な制約が考えられる.それはコミュニケーションの範囲を自分で制限できるということである.

一方,物理的制約がない Web 上ではコミュニケーションの範囲が規定できない.そこで,コミュニケーションの範囲を制限するために,Web ページ間の距離を定義する必要がある.Web 上での距離を基準としてコミュニケーションの範囲をコントロールし,

[†] 法政大学
Hosei University

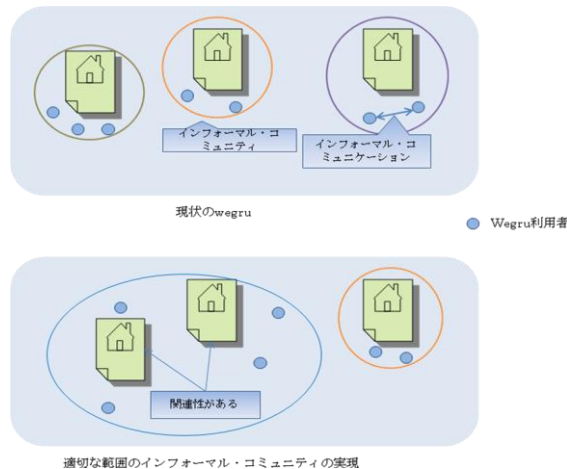


図 1 適切なインフォーマル・コミュニティ範囲の実現

ホームページをグループ化して、グループ内のユーザにコミュニケーションを行う「叫ぶ」機能を提案する[8]。

3. 距離の規定

3.1 距離の選定基準

「叫ぶ」機能を実現するにあたり、Web 世界における距離を規定する必要がある。そこで、現実世界の物理的制約としての空間の代替として、Web ページ間の内容や構成の関連性を考慮した距離(空間距離とする)と、物理的制約のない Web 世界の特徴を利用して、過去にその Web ページを閲覧したユーザへの時間的な距離(時間距離とする)の2種類の選定基準を挙げる。

3.2 空間距離を基準とした範囲

現在閲覧している Web ページの内容や、構造をもとに Web ページ間の距離を定義し、「叫ぶ」範囲を調整する、Web ページをベースにしたインフォーマル・コミュニティを形成する。今回は、内容の類似性を把握することができないため、ホームページの構造をもとに距離を定義する。

現在、多くのホームページが Web 上に存在しているが、多くの Web デザインマニュアルにおいて木構造が推奨されており、木構造に基づき研究されている[7]。よって今回

はホームページがトップページからコンテンツごとに徐々に詳細化された Web ページに枝分かれする構造を持つと仮定し、これを意味的な木構造と定義する。その構造の中の階層の定義を行い、「叫ぶ」範囲を調整する[8]。

3.3 時間距離を基準とした範囲

あるユーザが現在閲覧している Web ページを過去に訪れたことがあり、現在 wegru を使用している他のユーザを「叫ぶ」範囲とし、ユーザが Web ページを閲覧した時刻を過去に向かって遡り、距離を調整していく、ユーザをベースとしたインフォーマル・コミュニティを形成する。

ある Web ページを訪れたユーザが最後に閲覧した時刻を記録し、現在を基準として、現在と最後に閲覧した時刻の差を距離と定義し、その距離の幅で「叫ぶ」範囲を調整する。例として、「叫ぶ」範囲を現在から、一日以内、一週間以内、一ヶ月以内と過去に向かって広げていく。

4. 実際のホームページの構造

本稿では、空間距離における Web ページの構造を基準とした範囲の規定についての検討を行う。

3.2 節でホームページは木構造で構成されていると仮定したが、実際のホームページを検証すると、図 2 に示すように、意味的には木構造として構成されているが、つくりとして完璧な木構造で構成されているホームページは多くない。スタティックなホームページであっても、Web ページから様々な Web ページに向かってハイパーリンクが貼られており、その中から意味的な下位階層の Web ページを判断するのは難しい。

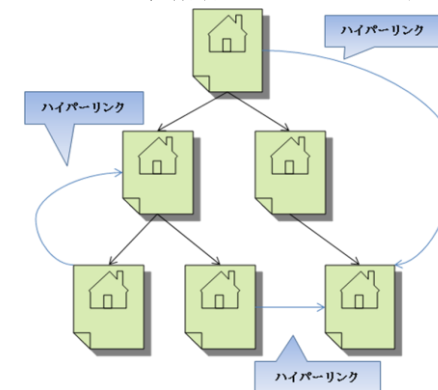


図 2 実際のホームページの構造

5. ホームページ群選択方式

本章では、ホームページの意味的な階層構造を判断するにあたり、ホームページの URL のドメインやディレクトリ階層から判断して下位階層の Web ページ抽出し、ホームページの意味的な階層構造を把握する方式の検討を行う。

5.1 仮説

- ホームページの構成において

この方式で下位階層の Web ページを取得できそうなホームページの構成として、スタティックな HTML で記述されたホームページがあげられる。

一方、この方式で下位階層の Web ページの取得が難しいと思われるホームページとして、全体的に、動的な Web ページを生成するホームページがあげられる。例えば、asp のみで構成されている Web ページは URL に階層がなく、ルートディレクトリのみ存在しているため、Web ページ内のすべてのハイパーリンクを抽出してしまうと予想される。また、Flash やフレームを使ったホームページでは意味的な下位階層のハイパーリンク先の Web ページを取得できるが、実際にハイパーリンク先の Web ページに移動しても URL は変化しないケースが存在する。

- ホームページの種類において

この方式で下位階層の Web ページを取得できるとされるホームページの種類として、コンテンツが頻繁に、また大幅に更新されないホームページがあげられる。

一方、この方式で下位階層の Web ページの取得が難しいと思われるホームページの種類として、ブログ等の流動的なホームページがあげられる。

ブログサイトは意味的においても、実際のつくりにおいても階層が存在しない。ほとんどのブログサイトのディレクトリ構成は `http://ドメイン/ユーザ名/index.html` となっており、ブログユーザは自身のユーザ名のついたディレクトリにコンテンツを追加し、すべて同ディレクトリの中に保存される構成となっている。よって今回の方式に限らず、ブログサイトにおいて構造的に距離を把握するのは困難であると予想される。

5.2 ホームページ群選択方式の検討

この方式では、まず着目した Web ページの URL を取得し、その Web ページがどのディレクトリ内に存在しているかを認識する。

次に、Web ページ内に存在するハイパーリンクをすべて検出し、そのリンク先を判定する。リンク先の Web ページが他のホームページである場合や、着目した Web ページよりも上位のディレクトリに含まれている場合はそのリンク先を弾くことで着目した Web ページと同じディレクトリ、もしくは下位のディレクトリにある Web ページのみ

を取得する。

図 3 で例をあげると、現在閲覧している Web ページが `http://www.hosei.ac.jp/hogaku/index.html` とすると、後ろから数えて 1 番目のスラッシュまでを削除し、Web ページが含まれているディレクトリを認識する。現在のディレクトリは `www.hosei.ac.jp` の中の `hogaku` となる。次に Web ページのハイパーリンクを検出し、同じディレクトリ内に存在しないハイパーリンクを削除する。つまり URL が `http://www.hosei.ac.jp/hogaku/` で始まるハイパーリンクのみを抽出する。

また、着目したページがドメインのみの場合は、サブドメインが空白もしくは、`www` の場合に限り、別のサブドメインのハイパーリンク先もその Web ページの下位階層と判断する。例えば、`http://www.hosei.ac.jp/` に `http://koganei.hosei.ac.jp/` といった Web ページへのハイパーリンクが存在した場合、この Web ページは着目した Web ページの下位階層と判定する。しかし、`http://www.hosei.ac.jp/hogaku/` のようにディレクトリへのパスが存在する URL の場合は `http://tama.hosei.ac.jp/hogaku/` へのハイパーリンクが存在しても、下位階層とは判断しない。

抽出された Web ページが、着目した Web ページにとって実際に意味的な下位階層の Web ページにあたるのか評価を行う。

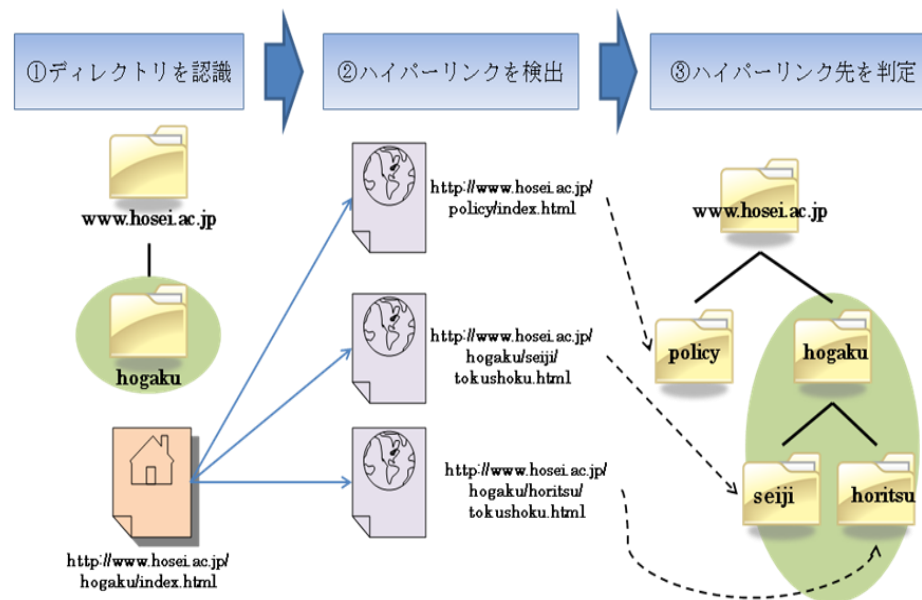


図 3 ホームページの URL を利用した選定方式

6. 調査

上記の方式で意味的な下位階層の Web ページが抽出可能なか以下の手順で調査を行う。

まず、着目した Web ページの意味的な下位階層である Web ページへのハイパーリンクの数を調べる。これを X とする。次に、今回の方式で実際に取得できた Web ページのハイパーリンクの数を調べる。これを Y とする。そして、 Y の中で意味的な下位階層である Web ページへのハイパーリンク数を調べる。これを Z とし、以下の式を定義する。

$$\blacksquare \quad \text{下位ページ含有率} = \frac{Z}{X} \times 100(\%)$$

$$\blacksquare \quad \text{下位ページ認識率} = \frac{Z}{Y} \times 100(\%)$$

下位ページ含有率は、本来の意味的な下位ページの中から今回の方式で、どれだけの割合で下位ページを取得できたのかを示す。下位ページ認識率は実際に取得した Web ページの中で、どれだけの割合で本来の意味的な下位ページを取得できたのかを示す。その様子を図 4 に表す。

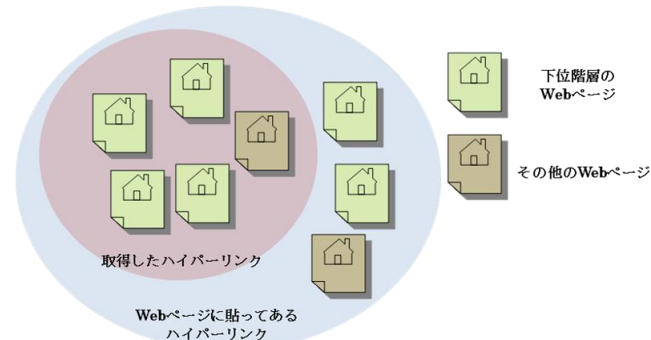
下位ページ含有率が低いということは、本来形成すべきインフォーマル・コミュニティの範囲に足りてないということである。しかし 0 パーセントでない限り、意味的な下位階層の Web ページの取得は可能であり、効果が薄くなってしまうが「叫ぶ」意味は存在するということになる。

一方、下位ページ認識率が低いということは、それだけ関連性のない Web ページとインフォーマル・コミュニティを形成するということである。

よって、今回の調査の判定では下位ページ認識率に基づいて下位階層の Web ページ取得の評価を行う。この数値は、高ければ高いほど理想的ではあるが、取得した Web ページの絶対数が少なければ、必然的に数値は低くなってしまう。よってそれを考慮し、今回は下位ページ認識率が 80%以上を適応領域とし、この方式において、意味的な下位階層の Web ページを取得できていると判断する。

実際の調査では、1 つのホームページのすべての Web ページを調査するのではなく、ホームページのつくりは統一していると想定し、図 5 に示すように 1 つのホームページにつき、木構造の 1 つの枝を調査の対象とする。その際、木構造の葉については調査の対象に含めない。その理由として、葉のページでは意味的な下位階層の Web ページが存在しないためである。

以上により、今回検討した方式において、どのようなホームページから意味的な下位階層の Web ページが取得できるのか、一般サイト（大学、病院、企業、公共等）、EC サイト、ブログサイト、ポータルサイトの分類ごとに調査した。



$$\text{下位ページ含有率} = \frac{\text{取得したハイパーリンク中の下位階層のWebページの数(4)}}{\text{下位階層のWebページの数(6)}} = 67\%$$

$$\text{下位ページ認識率} = \frac{\text{取得したハイパーリンク中の下位階層のWebページの数(4)}}{\text{取得したハイパーリンク中のWebページの数(5)}} = 80\%$$

図 4 下位ページ含有率と下位ページ認識率

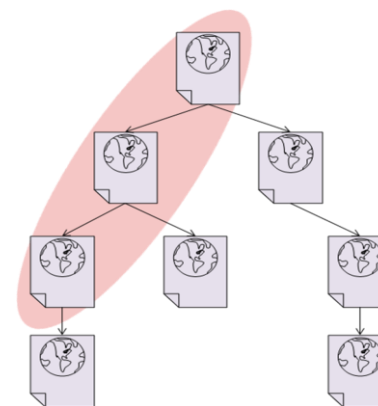


図 5 ホームページの調査する対象

7. 結果と分析

前章の手順で調査した結果得られたグラフを図 6 に示した。横軸が下位ページ認識率、縦軸が相対度数を表す。本稿で検討した方式の適応領域を明らかにするために、下位階層の Web ページを取得が出来るとする、下位ページ認識率が 80%以上の Web ページについて分析を行う。

- ホームページの構成において

ホームページの構成に基づいて分析を行うと、下位ページ認識率 80%以上の Web ページでは、スタティックな HTML を基本とするつくりの Web ページで占められていた。しかし、Flash や java script を含む Web ページがまったく存在しないというわけではなく、あくまでコンテンツの一部として組み込まれているものは取得可能である。

一方、CGM 等を主体とした、Web ページ全体として動的に構成されている Web ページは下位ページ認識率 80%以上のエリアには存在しない。

- ホームページの種類において

ホームページの種類に基づいて分析すると、下位階層の Web ページを取得できたとされる Web ページの特徴として大学や病院、市役所が挙げられる。この 3 つのホームページでは、調査した Web ページの約 9 割が下位ページ認識率 80%以上のエリアに存在している。

- ホームページの階層において

ホームページの階層に基づいて分析を行う。まず第 1 階層の Web ページ、つまりトップページの割合をみると、下位階層が取得できる Web ページは、調査した全ての第 1 階層の Web ページの 2 割しか存在しない。続いて、第 2 階層の Web ページは、調査した全第 2 階層の Web ページの約 5 割、第 3 階層は約 6 割、第 4 層は約 4 割の Web ページが下位階層を取得できる Web ページであることがわかった。

8. 考察

ホームページの構成に基づいた分析の結果、今回の方式では、スタティックな HTML で記述された Web ページでのみで使用でき、ダイナミックに構成されている Web ページでは下位階層を取得出来ないという仮説が証明された。そこで、動的な Web ページを生成するホームページを、Flash を使用している Yahoo!天気情報[9]を例になぜ不可能であるか考察する。

Yahoo!天気情報では、トップページに日本地図が Flash で表示されており、地図上

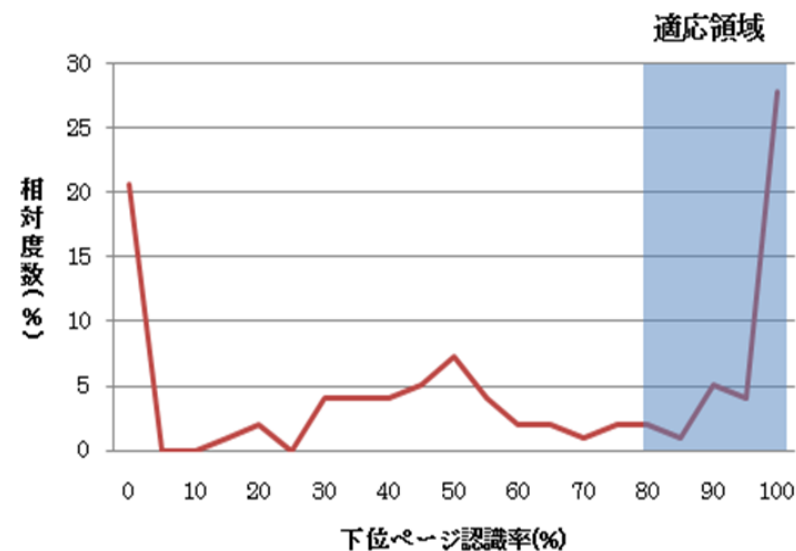


図 6 下位ページ認識率の相対度数グラフ

をクリックして範囲を狭めていくつくりになっている。しかし、範囲を狭くし、意味的に下位の Web ページへ移動したにもかかわらず URL は変化しない。つまり、日本全体を見ている時も、東京を見ている時も、ハイパーリンクを抽出すべき URL が同じなので、取得数つまり分母は変わらず、分子だけが減少し、適応領域に達することが難しくなる。

上記の他にも、データベースと連動して動的に生成される Web ページ等は、ディレクトリによって管理されていないため、この方式では下位階層の Web ページを抽出できない Web ページも存在した。

次に、ホームページの種類について考察する。コンテンツが流動的でないホームページであれば、取得が可能であると仮説を立てた。しかし、先の分析で得られた大学、病院、市役所以外にも、例えば企業サイト等も下位階層を抽出されると予想したが、芳しい結果は得られなかった。その理由は、下位階層の重複であると考えられる。下位階層の重複とは、意味的に 2 つ以上の上位階層に属している Web ページが存在することである。例を挙げると、企業の研究所についての Web ページに移動するためには、トップページ→研究ページ→研究所ページやトップページ→施設ページ→研究所ページといった意味的な経路が複数あるが、実際のつくりでは 1 つのディレクトリにのみ、その Web ページ

は存在している。研究ページのあるディレクトリに属しているとすれば、施設ページにおいて今回の方式で下位階層を抽出すると、ハイパーリンクを取得できないため、下位ページ認識率は低くなってしまふ。大学や、病院等はそれぞれのカテゴリやコンテンツの独立性が高いため、下位階層の Web ページが重複せず、高い下位ページ認識率が出されたと思われる。他にも EC サイトのカテゴリごとに分類されるページも独立性が高いため取得できる。

ホームページの階層についての分析の結果、第 1 階層つまり、トップページでは下位階層を抽出しづらいということが判明した。その理由として、この方式では現在着目している Web ページがあるディレクトリより、下位の Web ページを取得するため、トップページでは、他ドメイン以外のすべての Web ページを抽出してしまうからである。

また、実際に wegru において「叫ぶ」機能を使用する際に、今回の方式を適用するときは、まずソースや URL、ハイパーリンクを読み取り、現在閲覧している Web ページがディレクトリで管理されているか、Flash 等が使われているか、どんなホームページか（大学ならばドメインが ac）を先に検証し、ユーザに「叫ぶ」ことが出来るかどうかを通知する機能が必要になってくる。

9. おわりに

今回検討した方式の適応領域は、トップページを除き、スタティックに構成されており、意味的に下位階層へ移動するにつれて詳細化される時、カテゴリやコンテンツの独立性が高い Web ページとなる。

本稿では Web ページ間で適切なインフォーマル・コミュニティを形成するにあたり、ホームページ内の意味的な構造を把握するために、URL を利用したホームページ群の選択方式を検討し、評価を行い、その適応領域を明らかにした。

今後の課題として、動的に生成されていく Web ページにおいて構造関係を把握するための方式を検討し、また、ブログ等の構造を持たないホームページの距離や、内容による Web ページ間の距離を規定する必要がある。さらに「叫ぶ」機能の実装方式や、実際に運用する際のコミュニケーション形態を追求していく。

参考文献

- 1) Kraut, R. E., Fish, R. S., Root, R. W., and Chalfonte, B. L.: Informal communication in organizations: Form, function, and technology, In S. Oskamp & S. Spacapan(Eds.), Peoples's Reactions to Technology, pp.145-199, Sage Publications London,1990.
- 2) 後藤慎弥,金井敦: Web 上でのインフォーマル・コミュニケーション実現方式の提案,第 74 回グループウェアとネットワークサービス研究発表会
- 3) SPEEKY, <http://speaky.jp>

4) Nesutu, <http://adgjm.net/nesutu.htm>

5) Saul Greenberg,Michael Rounding. "The notification collage: posting information to public and personal displays". SIGCHI,pp.515-521,2001.

6) 西本一志: インフォーマル・コミュニケーションによる知識共創場の構築, 共創と複雑系シンポジウム予稿集,計測自動制御学会システムインテグレーション部門共創システム部会, pp.17-26,2006.

7) 熊谷知久,寺田実: Web サイト改良支援のための構造と履歴の視覚化システム, <http://www.interaction-ipsj.org/archives/paper2002/pdf2002/122paper.pdf>

8)山口亮輔,金井敦: Web 上での「叫ぶ」範囲選択方式の検討, 第 18 回 マルチメディア通信と分散処理ワークショップ (DPSWS2010)

9)Yahoo!天気情報, <http://weather.yahoo.co.jp/weather/>