

パーソナルサーバを用いた個人向け情報提供システムの検討

高山 貴行[†] 沼田 靖昌[†] 湯山 一郎[†]

本稿では、WWWにおける情報提供、及び個人から提供された情報をパーソナルサーバに蓄積し、その情報群からユーザの興味に応じた情報を選択、提供する個人向け情報提供システムの構築手法を提案する。本稿では、ニュース情報に対して万人向けにカテゴリ化した場合の性質を考察し、個人の興味にあった情報を提供するために必要であるとの結論を得た。そのため、ユーザが自らカテゴリを作成、選択する事によって、自分の興味に的確に合うカテゴリ群を生成する手法を提案する。これにより、本システムは、個人の興味に合わせた情報の提供を行う事が出来る。さらに、情報が提供されてからの経過時間に着目し、検討を行った。その結果、ユーザによるカテゴリ化の有効性、特に、ユーザの生活サイクルに合わせた情報優先度の変化を与える事が有効であることを確認した。

A study on a personal content navigation system using personalsever

TAKAYUKI TAKAYAMA,[†] YASUMASA NUMATA[†] and ICHIRO YUYAMA[†]

We propose a method of construction for personal content navigation system. The system accumulates the World-Wide-Web informations into the personal server. And the server selects and presents the information according to the interest of the individual user.

In order to present the information united with an individuals interest, it is necessary to prepare the category groups which absorbed the difference of interest of each individuals. The appropriate category groups for the individual user is produced by the user's action of build-up and select of categories. And, the system change the priority of information with the lapsed time since information was presented. Three kinds of operations using lapse time was evaluated. As a result, it is shown that to changed priority according to user's lifestyle habit is effective.

1. はじめに

近年インターネットやTVなどの情報媒体が多く存在し、発信される情報は増加の一途をたどっている。今後その傾向はますます強まる。しかしながら、個人の情報受容能力には限界があるため情報の選別が必要となり、その作業量は発信される情報量が増えるほど増していく。その結果、ユーザは実際に欲しいと思える情報を短時間で受け取るのが難しくなった。そこで情報の選別にかかる作業量を少なくし、個人の興味に合わせた情報を提供する手法の開発が必要となった。

一方で、情報媒体のデジタル化に伴い、情報検索技術が発展し、同一の情報量の中から必要な情報を探し出すことは以前に比べ容易になってきている。しかし、そのほとんどのシステムは万人向けの情報提供を行っているため、個人の興味が多様化している現在、ある一定以上の作業量の削減は難しい。このような作業量

を可能な限り削減したいという要求に対して、様々なパーソナライズサービスが提案されている³⁾。

本研究ではユーザの興味に合わせたカテゴリ分けを取り入れる事で、よりユーザの興味に合致した情報を容易に提供することを目指している。本研究では、提供する情報としてテキストを用いたニュース情報に絞っている。ニュース情報をどのようにカテゴリ分けしていけば個人の興味に合った情報を提供できるのかを考察し、それに沿ったカテゴリ群を用意する手法を提案する。また、比較的容易にユーザの趣向が反映されるように考慮したツールの開発を行う。

2. 情報提供カテゴリの選定手法の検討

本節では万人向けの情報提供用カテゴリの性質を考察し、構築と評価を行っていくことで、個人向けの情報提供にふさわしいカテゴリの制定を行うための条件を検討した。

2.1 万人向けカテゴリの性質

インターネット上の一般的なニュースサイトにおけるカテゴリ分類は、万人にとって情報の閲覧がしやす

[†] 宇都宮大学大学院工学研究科
Graduate School of Engineering, Utsunomiya University

表 1 アンケート結果

挙げられたカテゴリ名の候補数	40
適切と思われる分類カテゴリの平均	5.81
適切と思われる分類カテゴリ数の標準偏差	2.73

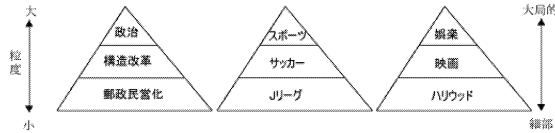


図 1 階層的に示されたニュース情報の例

い構造になっている。そのための性質として以下の点が挙げられる。

- カテゴリ分けを行った時に、可能な限り全体の情報量を損なわない。
- カテゴリの数が適切である。
- そのカテゴリが何を示すかがカテゴリ名から判断できる。
- 各カテゴリの粒度 (2.2 節参照) が適切である。

以上の点を満たしているときに、万人向けの情報提供カテゴリが生成されていると考えられる。

今回、ニュースのカテゴリ群の制定を行うにあたって、38名(19~25歳)にニュースを分類するカテゴリのカテゴリ名としてふさわしいもの、及びニュースを分類するカテゴリ数はどの程度が適当であるか、というアンケート調査を行った。その結果を表 1 に示す。ここでは言葉の近さを表す相関度 (2.3 節参照) という尺度を用いて制定手法を検討した。

2.2 情報の粒度

情報の粒度とは、情報がどの程度の大雑把さを持つかの指標である。ニュース情報に関して階層的に構築した例を見てみると、図 1 のように大局的なカテゴリから細部なカテゴリまで様々な大きさのカテゴリが存在する構造となる。実際には各カテゴリ間でつながりが生じているため、ニュース情報の各カテゴリを、綺麗に入れ子型の階層構造にする事は難しい。よって、カテゴリ毎といったような曖昧な分類による階層構造を構築する場合が多い。このそれぞれのカテゴリの持つ大雑把さが粒度と言えらる。大雑把さを具体的に数値に示すには、社会を反映したある母集団の中で、そのカテゴリを含むものがどの程度存在しているかという方法が考えられる。すなわち、二つのカテゴリが同程度のオーダーで、母集団中に存在するのであれば、粒度が等しいといえる。

2.3 相関度

相関度とは、言葉の近さを表すものである。ある母

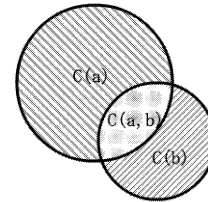


図 2 相関度

集団内のカテゴリ名候補 a の存在するページ数を $C(a)$ とし、同じ母集団内のカテゴリ名候補 a と b が共に存在するページ数を $C(a,b)$ とする。また、ある数値 m,n のより小さい値を取り出す関数を $\text{MIN}(m,n)$ とする。その時、カテゴリ名 a,b の相関度 $h(a,b)$ は

$$h(a,b) = \frac{C(a,b)}{\text{MIN}(C(a),C(b))} \times 100 \quad (1)$$

と定義する。今回は $C(a),C(a,b)$ の値として、検索エンジン google での検索結果数を用いた。

2.4 処理の流れと評価法

アンケートの結果から得られたカテゴリ群から、より万人向けのカテゴリの性質を有するようにカテゴリ群を作成する。そのために、相関度を用いて、以下に述べる手続きの概要に従い、不適切だと思われるカテゴリ名を削除して行く。実際の手続きに関しては一般に最適なものを見つけるのは難しく、本稿では概要に沿ったの手続きを検討し、その中から制定した。

アンケートでは、カテゴリ名候補として適切と思うものを推奨させているので、そのカテゴリが何を示すかはカテゴリ名から判断できる。それを踏まえた上で、以下の (a),(b),(c) 3つのフィルタリングを順番や適応する数を変化させて行うことで、初期の候補に挙げられたカテゴリ群から、万人向けの性質を持ったカテゴリ群をいくつか生成した。

- 各カテゴリの粒度を適切にするために、あるカテゴリ名の含む情報量が極めて大きいと思われるものを取り除く。例としては情報や社会といったカテゴリ名である。
- 各カテゴリの粒度を適切にするために、あるカテゴリ名の相関度が1つだけ飛びぬけた値を持つものを取り除く。例としてはサッカーというカテゴリ名に対してのサッカー日本代表というカテゴリ名である。
- 情報量を損なわずにカテゴリ数を減少させるために、あるカテゴリ名と他の各カテゴリ名の相関度の値がほぼ変わらないものを取り除く。このように横断的な値をとっているものは例として挙げると、おすすめ、トレンドと言った他の様々な単

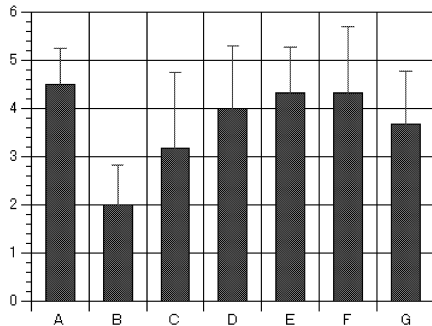


図3 フィルタリングとカテゴリ群の好ましきの変化

語と接続して意味を成すカテゴリ名である。

上記の処理で得た7つのカテゴリ群を各群にA~Gといったグループ名をつけた。また、それぞれのカテゴリ群が元の40個のカテゴリ群に比べてどれだけ好ましくなったか、を6人の男女(21~26歳)に7段階評価させた。値は0~6で0は元のデータと変わらない、または悪化したとし、数字が大きいほど好ましいとした。

2.5 結果

図3に結果を示したが、個々の評価に大きなバラつきが生じた。

これは個人の興味の差異から生まれるものと考えられる。従って、その差異を吸収したカテゴリ群を用意することが、個人の興味にあった情報を提供するために必須である。

3. 個人向け情報提供手法の構築

ここでは前節に示した情報提供における個人の興味の差異を吸収し、ユーザの興味に沿ったカテゴリ群の生成手法、及びそれを踏まえた上での個人向け情報提供システムの構築法について述べる。

3.1 個人向け情報提供用カテゴリ群の構築概念

前述したように、個人の興味に合わせた情報の提供には、個人の興味に合ったカテゴリ群を用意する必要がある。しかし、個々人の興味が大きく違うことは明白である。従って、システムの初期状態として個人の興味を充分反映したカテゴリ群を準備するのはできない。そこで、新たなカテゴリをユーザが自ら生成するという手法を用いてシステムを構築した。これによってユーザがどのような事に興味を持っていたとしても、ユーザは自分の興味に合った情報を的確に受け取ることができる。

3.2 実際のシステム構築手法

本システムでは、システムの初期状態として複数のカテゴリを設定し、容易にカテゴリの追加と編集を行

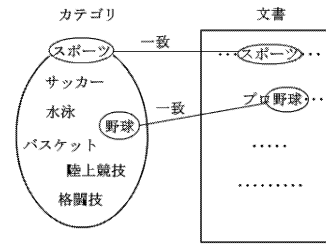


図4 メタデータ生成時の比較

えるように設計した。初期状態として準備したカテゴリは、万人向けの性質を持つ。これら複数のカテゴリを存在させたのは、ユーザが新規に作成するカテゴリ数を減少させ、ユーザの手間を省くという事を意図している。また、カテゴリの編集、追加は随時行うことを認め、ユーザの興味に変化が生まれた場合でも容易に対応可能とした。

この方法で作成されたカテゴリを含め、存在する全カテゴリを用いて提供する情報に対して全文検索を行い、その情報に対するメタデータを生成する。メタデータはその情報がどのカテゴリにどの程度属しているのかを、割合で示すデータとした。ある情報(x)の文中で、あるカテゴリ(m)のカテゴリ名、又はカテゴリの要素と一致(図4参照)した数を $find(x,m)$ とする。その場合、その情報(x)がカテゴリ(m)に対してどの程度属しているのかを $became(x,m)$ とし、以下で示す。

$$became(x,m) = \frac{find(x,m)}{\sum find(x,i)} \quad (2)$$

これにより、全体としては、

$$\sum became(x,i) = 1 \quad (3)$$

となるメタデータが生成される。このように割合を用いてメタデータを生成する事によって、あるカテゴリの構成要素数が他のカテゴリに比べて少ない場合でも、同列に評価を下すことができる。

情報の受け取り時には、ユーザは自ら作成したカテゴリも含め、存在するカテゴリの中から自分の興味のあるものを選択する。そして、対象情報のメタデータにユーザの選択したカテゴリが含まれた場合にその情報を提供する。

3.2.1 経過時間評価

ニュース情報の性質として、新たに提供されたものほど分かりやすく表示され、掲示順の上位にすることが望ましい。しかし、具体的にどのような関数を用いるのが適切なのかが不明である。そこで今回3種類の関数、 $time(A)$, $time(B)$, $time(C)$ を用意した。それぞ

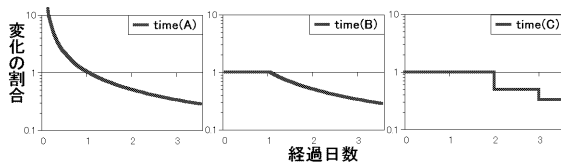


図 5 経過時間評価関数

れの評価値を図 5 に示す。

3.3 情報受け取り

対象情報をシステムがどのようにしてユーザの興味にあっていると判断するか。また、ユーザは対象情報をどのようにすれば受け取ることが出来るかを説明する。

システムはユーザの持つユーザ情報を元に興味のあるカテゴリと提供される各情報のメタデータ(その情報があるカテゴリに属している割合)を比較し、一致した部分を積算する。すなわち、各情報(x)がそのユーザにとってどれだけ必要であるか、をシステムが推測し情報提供優先度(need(x))とし、あるカテゴリ(m)がユーザの興味情報と一致したかどうかを判断する関数を match(m) とすると、

$$match(m) = \begin{cases} 1 \dots m \text{ がユーザ情報と一致するカテゴリ} \\ 0 \dots m \text{ がユーザ情報と一致しないカテゴリ} \end{cases} \quad (4)$$

$$need(x) = \sum \{ match(i) \times became(x, i) \times time \} \quad (5)$$

と表す。情報受け取りの際には need(x) を降順でソートして、ある閾値以上であればタイトル一覧を表示する。ユーザは、その一覧を参照する事で詳細情報を見る事ができる。

3.4 全体の構成

今回開発した個人向け情報提供システムは、HTML 文書と Perl を用いて書かれた CGI で構成している。よって、本システムはインターネットブラウザを用いて使用することができる。実際の使用画面(図 6)には HTML のフレーム機能を用いて画面左側にメニューが表示され、右側に内容が表示されるようになっている。

4. 実験

4.1 実験方法

実験の目的は、ユーザが自らカテゴリの追加と編集を行うことで個人向けカテゴリをを生成でき、個人の興味に合った情報を提供する事を確認すること。また、情報提供からの経過時間によってユーザの行動が



図 6 実際のシステム外見

どのように変化するかを調べることである。そのために、提供される情報量を一定量に保つため、深夜帯~朝(AM0:00~AM8:00)を除いて1,2時間毎にニュース提供サイトで新規に発信されているニュースを本システムの情報発信を用いて提供した。ニュース提供サイトとしては、google ,yahoojapan ,nhk オンラインを用いた。被験者は各 time 関数に 11 名を割り当て、全体の被験者は 20~24 才の情報工学専攻の学生計 33 名とした。「情報受け取り」の閾値は 0,2,初期のカテゴリは政治、経済、スポーツ、サイエンス、エンターテインメント、事件事故の計 6 カテゴリを用意した。

評価のために各ユーザのアクセス状況や、どの情報を選択したかをログとして記録した。また主観的な評価を得るため、システム使用期間の後に同一サイト上でアンケートを行った。

客観的な評価の尺度として、第 2 章で説明したシステムの推測したユーザの興味のある情報と実際のユーザの参照行動がどれだけ一致したかを示す適合率を用いる。情報受け取り時にはタイトルが一覧表示される。その際上位にあるものほどシステムの提供優先度が高い情報である。この順番に一番上から興味順位をつけていく。ユーザ(u)が興味順位 r 番目の情報を参照した時、その時提供された全情報の中でどのくらい上の位置にあるかを示す情報の順位割合 rate(u,r) は、

$$rate(u, r) = \frac{r}{\text{その時提供された全情報数}} \quad (6)$$

となる。その時、適合率 adapt(u,r) は以下のように表す事が出来る。

$$adapt(u, r) = \frac{0 \sim rate(u, r) \text{ までの参照回数}}{\text{総参照回数}} \quad (7)$$

4.2 実験結果と考察

実際に本システムを 10 日間使用した結果では、総情報提供数は 923, 期間内の情報参照回数は延べ 815

回であった。また記事参照回数が 20 回以上のユーザが 17 名であった。さらに、アンケートによる評価を行い、time(A)10 名、time(B)10 名、time(C) 7 名、計 27 名の回答を得た。

図 7 に今回の実験の適合率を示す。ここでの適合率はそれぞれのユーザの適合率の平均をとったものである。図 8 では情報の提供優先度に反映する 3 種類の time 関数それぞれの結果を示してある。また、経過日数による情報提供量の変移を各 time 関数毎にまとめたものを図 9 に示す。

まず、図 8 から今回の実験では提供日数に伴う情報提供量がある程度一定量に抑えられている事が見て取れる。これによって、長期間使用したとした場合でも適合率の低下は無くなる。図 7 から今回の実験では一日経過していない時に提供優先度を増幅し、その後連続的な減衰を行う time(A) が最も良い結果となった。また 1 日経過した後に連続的に減衰して行く time(B) と、1 日経過した後に経過日数を整数値とし、離散的に減衰していく time(C) は、ほぼ同様の値を示している。ユーザの主観評価であるアンケート(図 9)の結果を見ると time(A) を用いた人の 5 割が十分に満足した評価を 8 割が好意的な評価を下し最も良い評価を得ている。Time(B), time(C) では非常に満足した人がいなかったが、6~7 割の人は比較的好意的な評価を下している。

このことから、興味のある情報の中でもユーザが自らの生活サイクルにおいて他の情報提供システムや、各メディアから同等の情報を見聞する可能性が低いものが上位に来る事が望ましいという事が推測できる。よって、丸 1 日以上経過された情報はその間に一度またはそれ以上の回数見聞した可能性が極めて高く、その後提供優先度の連続的な減衰と離散的な減衰との間に大きな差異が出てこないと考えられる。

次に、各種アンケート結果を示す。

図 10 の「カテゴリの編集を行ったか」から、全体の 6 割程度の人がカテゴリの編集を行っていることが判った。表 2 に実際の実験前後のカテゴリの変化を示す。図 11 にカテゴリの編集を行わなかった理由を示す。ここから、編集を行わなかった人は初期に設定されていたカテゴリで十分であると答える人が多かったとわかる。この事から、初期に用意した万人向けのカテゴリで不足を感じた人は個人の興味に合わせてカテゴリを作成したと考えられる。しかしながら、アンケート自由記入欄では、作成したカテゴリは個人の興味を合わせたものなので、他のカテゴリに対して、より上位に位置して欲しいという意見もあった。これは

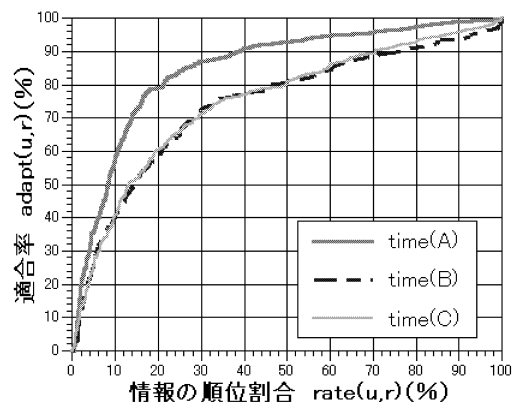


図 7 適合率

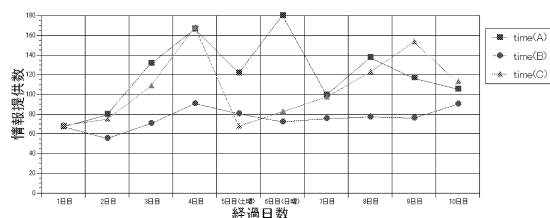


図 8 提供日数に伴う情報提供量の推移

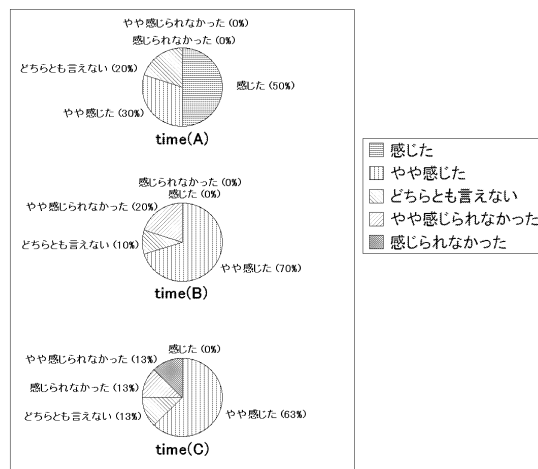


図 9 個人の興味に合った情報が提供されていたか(アンケート)

情報の粒度が小さいカテゴリは、より個人の趣向を反映していると考えられるからである。よって、情報提供優先度の設定をもっと細かくする必要がある。

「カテゴリ編集は簡単に行うことが出来たか」(図 12)を見ると、全体の 8 割以上がカテゴリの編集は容易であったと答えている。よって、ユーザにとって過大な負荷となることは少ないと推測される。しかし、アンケートにてフィードバックをかけた際の待ち時間が長かったとの指摘があった。蓄積されている情報が増加していくほどシステムに対しての負荷が大きくな

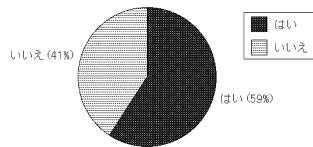


図 10 カテゴリー編集を行ったか (アンケート)

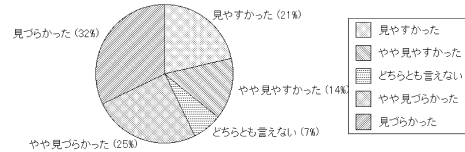


図 14 提供された情報は見やすかったか (アンケート)

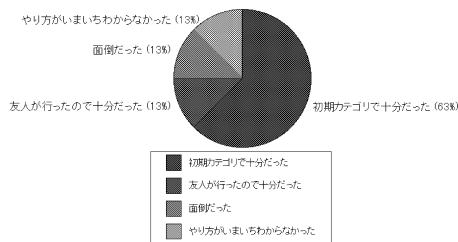


図 11 カテゴリー編集を行わなかった理由

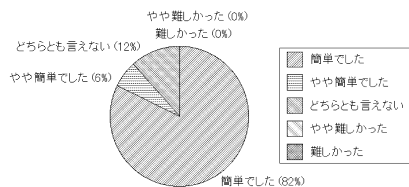


図 12 カテゴリー編集は簡単に行えたか (アンケート)

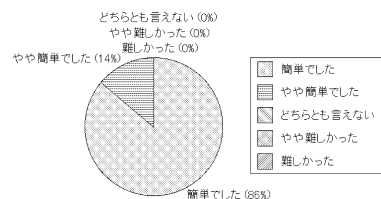


図 13 情報受け取りは簡単に行えたか (アンケート)

るといのは当然である。システム負荷については今後検討する必要がある。

「情報を容易に受け取ることが出来たか」(図 13)を見ると、全体の 86% が簡単だったと答え、また残りの 14% の人もやや簡単だったという好意的な評価を下している。これから見ても操作が容易であった事がわかる。

提供された情報は見やすかったか(図 14)については好意的な答えが 35%、否定的な答えが 57%と否定的な意見が多かった。これは本システムにおいて提供される情報がカテゴリを明示しないで優先順位順に並べた一覧表示であった事、参照したページが文字のみで表示され、単色であった事が大きな原因であると考えられる。

表 2 実験の前後におけるカテゴリの変化

初期カテゴリ	実験終了後のカテゴリ
政治	政治
経済	経済
スポーツ	スポーツ
サイエンス	サイエンス
エンターテイメント	エンターテイメント
事件事故	事件事故
	サッカー
	F1
	格闘技
	気象
	IT
	自動車
	国際
	宇都宮

5. 終わりに

本研究ではユーザによるカテゴリライズ及び提供からの経過時間からの評価を行うことで、個人の興味に合わせた情報が提供されることを確認した。また、情報の提供からの経過時間を用いての提供優先度は生活サイクルに合わせて変化させることが、個人の興味のある情報を提供するのに望ましいこともわかった。一方、カテゴリの編集を行ったものは、より上位に位置して欲しいといった意見が存在した。それ以外にも情報の見づらさなどのインターフェース面などの問題、システムの負荷に関する問題などが本システムには存在する。今後はその問題一つ一つを解消し、より良いシステムの構築を目指す。

参考文献

- 1) 北 研二他/”情報検索アルゴリズム”:共立出版；ISBN: 4320120361；(2002/01)
- 2) 徳永 健伸”言語と計算 - 5 情報検索と言語処理”:東京大学出版会；ISBN: 4130654055；5 巻(1999/11)
- 3) 橘高, 佐藤他/”パーソナライズ情報提供方式の提案と評価, 情報処理学会論文誌, Vol. 40 Num. 1 pp.175-187 (1999.01)
- 4) google : ”http://www.google.co.jp/”
- 5) yahoojapan : ”http://www.yahoo.co.jp/”
- 6) NHK オンライン : ”http://www.nhk.or.jp/”