

遠隔非同期 Web ユーザビリティ評価における 評価者とユーザの差異

中道上^{†1} 山田俊哉^{†2}
木浦幹雄^{†3} 栗山進^{†4} 上野秀剛^{†5}

本研究では遠隔非同期 Web ユーザビリティ評価に向けて、ユーザとのやりとりが一切無い状態で、ユーザの主観評価とインタラクションデータのみを用いた評価者の評価の差異について調べた。ユーザの評価結果と評価者の評価結果を比較、分析した結果、評価者がユーザと同様の評価をつけることは難しいことがわかった。しかし、ユーザビリティ評価の専門家であれば被験者の評価結果と相関があり、過大評価していないことがわかった。

Difference between Evaluators and Users in Remote Asynchronous Web Usability Evaluation

NOBORU NAKAMICHI^{†1} TOSHIYA YAMADA^{†2}
MIKIO KIURA^{†3} SUSUMU KURIYAMA^{†4}
HIDETAKE UWANO^{†5}

Recently, Some tools and environment may have applicability to remote asynchronous web usability evaluation are being put into place. We experimented for comparing between users' evaluation and evaluators' evaluation. At this time, evaluators don't see users' interaction in the experiment for recording users' interaction data. From comparison result, it is difficult to evaluate same evaluation for evaluators. But a usability expert in evaluators has a correlation with users' evaluation. And his evaluation only is not over-evaluating.

1. はじめに

Web サイトは企業の成長過程においてグローバル化が必要になった際の大きな役割を担っている。Web サイトの価値はますます高まってきており、現在は数百億円を超えるサイトもあると試算されている[1]。そのため Web ユーザビリティは、企業の売り上げに影響を与えるなど重要性が高い。

グローバル化を進めるにあたり、国内マーケティングの延長として進める場合が多く、現地に合わせた地域マーケティングが進められていない[2]。そのため Web サイトでは、言語のみを翻訳した海外向け Web サイトが多く見られる。しかし、市場の経済発展が進めば進むほど地域マーケティングの重要性が増し、Web サイトも地域に合わせて改善を進める必要がある。

Web サイトの Web ユーザビリティを改善するためにユーザビリティテストが実施されている。ユーザビリティテストは実際にユーザに操作してもらい、ユーザ自身に評価してもらう方法であり、ユーザトラブルを引き起こす問題や評価者の思いもよらない問題点を発見できるため広く用いられている。しかし実施するにあたってユーザを集めるためのリクルーティングコストが大きい問題が挙げられる[3][4]。特にグローバル化を進めるにあたって地域の人々からユーザを抽出し実施することは困難である。ユーザを集めることなく、ユーザが日常と同じ環境のもと、遠隔非同期でユーザビリティテストを実施することが可能であれば海外市場に向けた Web サイトの改善とともにコストの削減が期待される。

現在、Web カメラによる視線計測環境[5]や視線とマウス操作、Web 閲覧履歴を記録可能な Firefox アドオン[6]を利用することにより、遠隔非同期でユーザの閲覧状況を記録可能な環境が整いつつある。

本研究では遠隔非同期 Web ユーザビリティ評価の実現に向けて、評価者とユーザとのやりとりが一切無い状態で、ユーザの主観評価と評価者の評価の差異について分析を行った。評価者は評価する際、ユーザの視線やマウスの動きといったインタラクションデータのみを用いて評価した。

^{†1} 南山大学

Nanzan University

^{†2} 総合研究大学院大学

Hitachi Ltd.

^{†3} キヤノン株式会社

Canon Inc.

^{†4} 株式会社ミツエーリンクス

Mitsue-Links Co., Ltd.

^{†5} 奈良工業高等専門学校

Nara National College of Technology

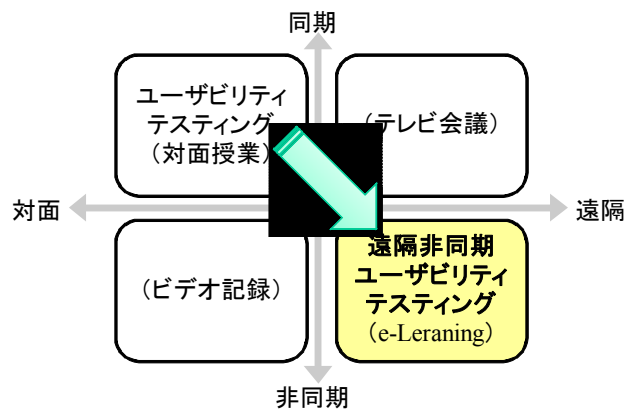


図 1 遠隔非同期ユーザビリティ評価の位置づけ
Figure 1 Remote Asynchronous Web Usability Evaluation Positioning

2. 遠隔非同期 Web ユーザビリティ評価

本研究における遠隔非同期 Web ユーザビリティ評価について定義する。図 1 に本研究における遠隔非同期ユーザビリティテストを示す。Web ユーザビリティ評価には、専門家が知識と経験から評価するユーザビリティインスペクションも行われているがユーザが直面する問題をうまく予測できないため、本研究では実際のユーザのインタラクションを観察するユーザビリティテストを対象とする。

遠隔とは対面でない実施状況を指す。実際のユーザビリティテストでは同じ空間のもと対面で実施されるため、ユーザの様々なインタラクションを評価者が観察することが可能である。そのうえで評価者はインタビューを実施し、ユーザビリティを評価することが可能である。

非同期とは同じ時刻に実施しない状況を指す。実際にユーザと評価者が同期した同じ時刻にユーザビリティテストを実施した場合、ユーザが問題に直面した時刻を評価者も共有することが可能である。そのためユーザが直面した問題を答えやすい、また評価者も問題箇所をユーザに確認することが容易である。

対面同期でユーザビリティテストを実施する場合には、ユーザと評価者がユーザビリティラボといった実施場所に集まる必要がある。しかし、遠隔非同期であ

ばユーザと評価者が同一の場所、時刻に実施する必要はなく、ユーザを集めるといったコストが少なくなる。また海外市場に向けた Web サイトにおいても対象となるユーザに対してユーザビリティテストを実施することが可能となり、地域に合わせた Web サイト戦略を進める大きな助けになると期待される。

3. 関連研究

ユーザビリティテストでは、一般的にユーザのインタラクションの様子を記録するためのカメラが設置されており、観察者はその映像を観察・記録する。ユーザのインタラクションをより詳細に調べるため、近年様々なツールが提案されており、様々なインタラクションデータを定量的に記録することが可能となりつつある。特に分析に利用されるインタラクションデータとして新たに視線の動きが利用され、ユーザの詳細なインタラクションの分析が可能である。

定量的なインタラクションデータを分析することにより、Web サイトおよび Web ページに存在するユーザビリティに関する問題点をより客観的に発見することができるようになる。Arroyo ら[7]は、マウスの動きをスクロール、読み、停止、メニュー、ランダムに分類し、これらの分類に基づいた分析を行うことで問題の箇所を指摘できる可能性があることを示唆している。また、Josephson ら[8]は、視線の軌跡を利用し、その軌跡のパターンから Web ページのデザインの問題点を指摘できる可能性があることを示している。このように Arroyo らや Josephson らは、ユーザビリティの評価者がユーザのインタラクションをより詳細に分析することを可能にした。しかし、遠隔非同期にユーザのインタラクションを分析した場合にも対面で実施した場合と同様の結果が得られるかについては実験されていない。

遠隔非同期なユーザビリティテストを実施するにあたって利用可能なツールとして Webjig[9]や ITR-Recorder/ITR-Player (Web InTeRaction-Recorder/Player) [6]がある。Webjig は Web サイトを利用するユーザのマウスカーソルの動きやクリック、キーボードの操作を記録すると同時に、Web ページの表示が変化することに Web ページの DOM (Document Object Model) を解析することで、ユーザ画面 (ユーザが利用している Web ブラウザに表示されている内容) を記録する。また、分析作業を支援するために記録した操作・表示内容を可視化、再生する機能を持つ。Webjig はクライアントサーバ型のシステムとなっており、Web サイトにアクセスしたユーザの情報を収集するシステム (Webjig::Fetch) と Web サイト評価者にユーザの情報を表示するシステム (Webjig::Analysis) の 2 つのサブシステムと、ユーザ行動情報を保存しておくための Webjig::DB から構成されている。しかし、視線データは記録していないため、マウス停止時のユーザのインタラクションを分析することが難しい。

ITR-Recorder は Firefox の機能拡張として動作し Web 閲覧時のユーザ行動を記録す

るためのシステムである。ITR-Recorder はユーザが閲覧している Web ページの内容とユーザ行動と同時に、ユーザの注視点の位置情報、瞳孔径を Eye Tracking Device から取得する。なお、Eye Tracking Device として nac 社製 EMR-AT VOXER が想定されている。ITR-Player は、ITR-Recorder で取得・保存されたユーザ行動データを利用して、ユーザ行動を可視化するためのシステムである。ITR-Player は同一の Web ページにおける複数のユーザ行動を比較分析するために、Collaborative Visualization を用いる。Collaborative Visualization は Web 閲覧時の個々のユーザ行動を可視化するとともに、同一の Web ページにおける複数のユーザ行動データを協調的に可視化することも可能である。

本研究では ITR-Recorder/ITR-Player を利用し、ユーザのマウスや視線の動きから評価者が対面で実施した場合と同様にユーザビリティ評価可能であるか実験、分析を行う。

4. ユーザと評価者による評価記録実験

まずユーザが閲覧した Web ページにおけるインタラクションデータを記録する実験を行い、ユーザのマウスや視線の動きといったインタラクションデータを記録した。また記録したインタラクションデータを再生しながらユーザが閲覧した Web ページ 51PV ごとの評価を記録した。遠隔非同期な状況を想定し、ユーザとのやりとりが一切無い状態で、評価者には記録したユーザのインタラクションデータを再生しながらユーザが閲覧した Web ページを評価した。

また本論文では、Web ページのカウント方法として PageView (PV) を採用する。PV とは Web ページのアクセスカウントの 1 つであり、ブラウザによって 1 回リクエストされサーバから読み出される Web ページを 1PV とカウントする。

4.1 実験環境

本研究に用いたインタラクションデータの記録についてと、ユーザによる主観的な Web ユーザビリティの評価結果を調べるためにユーザビリティテストの実験を行った。

本研究で用いた実験環境は以下のとおりである。

- ディスプレイ：液晶 21 インチ(有効表示領域：縦 30cm, 横 40cm, 解像度：1024 × 768pixel)
- 顔とディスプレイの距離：約 50cm
- 視線計測装置：NAC 社製 EMR-NC (視野角：0.28 度, 画面上の分解能：約 2.4mm)
- インタラクションデータの記録：ITR-Recorder (サンプリングレート：毎秒 10 回)
- インタラクションデータの再生：ITR-Player

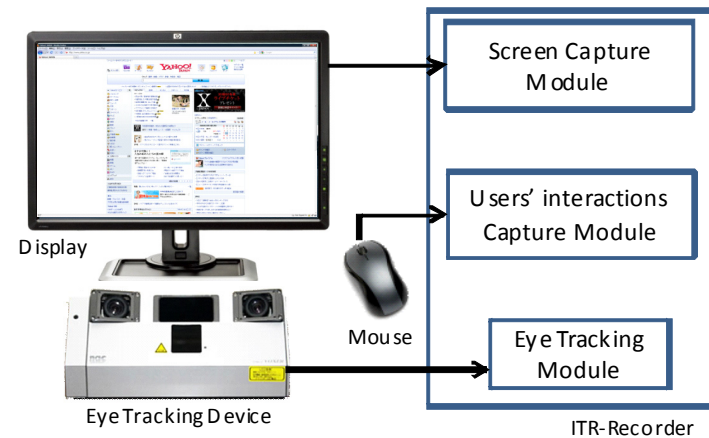


図 2 ITR-Recorder のシステム構成
Figure 2 System Architecture of ITR-Recorder.

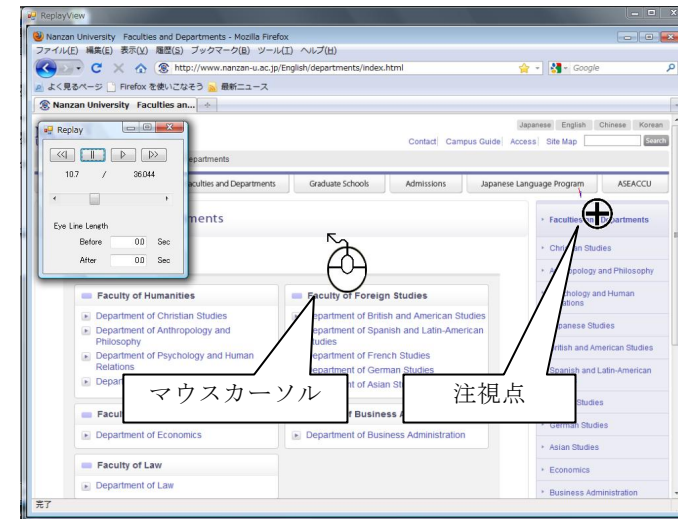


図 3 ITR-Player による再生画面例
Figure 3 Replay Function of ITR-Player.

4.1.1 インタラクションデータの記録 ITR-Recorder

ITR-Recorder (Web InTeRaction-Recorder)はユーザの日常と変わらない Web 閲覧時のインタラクションデータを記録するツールである。そのため、Web ブラウザ Firefox のアドオンツールとして開発されている。ITR-Recorder は、図 2 のように、Screen Capture Module, Users' interactions Capture Module, Eye Tracking Module Module の 3 つのモジュールから構成されている。そのためインタラクションデータとしてユーザが使用している Web ブラウザ (Firefox) に表示されている内容の画像ファイル、ユーザ操作 (マウス移動を知るためのマウスカーソルの位置座標、マウスクリック、ページスクロールを知るためのホイール回転量、ページ遷移履歴、タブ操作)、をユーザの視線の動きを知るための注視点の位置座標が記録可能である。ここで注視点とはユーザの視線と画面上の交点である。

4.1.2 インタラクションデータの再生 : ITR-Player

本実験では ITR-Player の Replay 機能を利用し、ユーザは自身のインタラクションの様子を確認した。ITR-Player は、ITR-Recorder で取得・保存されたインタラクションデータを利用して、ユーザのインタラクションを可視化するための独立したツールである。ITR-Player では個々のユーザのインタラクション分析を支援するための 3 つの機能(Summary, Replay, TraceLog)と、複数のユーザ間のインタラクションを比較分析するための Collaborative Visualization の 4 つの機能を提供している。Replay 機能は記録された Web ページの画像とインタラクションデータを統合して表示する。Web ブラウザの操作の再生時には、Web ページの画面上にマウスと注視点の座標にそれぞれアイコンを表示する。マウスクリックとホイール操作は、画面上にアイコンでイベントの発生を表示する。ITR-Player による再生画面を図 3 に示す。

4.1.3 実験手順

ユーザとなる被験者は、日常的に Web を利用している理工系の大学生 9 名である。被験者に、実在する 5 つ企業の Web サイトから大学卒業生の初任給を探すタスクを行うよう依頼した。タスクの数は 5 つの企業の Web サイトからランダムに選択した 1 タスクである。それぞれのタスクに対して以下の手順でユーザビリティテストを行い、1PV 毎にインタラクションデータを記録した。また、被験者の主観的な評価結果を調べた。実験手順を以下に示す。

手順 1. 初期設定として、被験者のディスプレイに各企業のトップページへのリンクを張った実験用 Web ページを表示しておき、タスクを実行するために被験者がそのリンクをクリックした時点から実験を開始する。

手順 2. ITR-Recorder を用いて、被験者のタスク実行中のブラウザ操作の様子を記録する。その際、評価者が被験者に対して質問するといったタスクの中断

につながることは行わなかった。タスクは被験者が初任給を見つけることができたときと申告した時点で記録を終了する。

手順 3. タスク終了後、被験者が訪れた Web ページを ITR-Player を用いて閲覧しながら、Web ページごとの使いやすさを下記の 4 つから選択するよう依頼する。

1. 使いにくい
2. どちらかといえば使いにくい
3. どちらかといえば使いやすい
4. 使いやすい

実験の結果、被験者が閲覧した Web ページのうち 51PV において被験者のインタラクションデータを記録し、アンケート結果を収集した。

4.2 評価者による評価

評価者となる被験者はユーザが閲覧した Web ページにおけるユーザの視線やマウス操作といったインタラクションを ITR-Player を用いて再生し、各ページの使いやすさを評価した。評価者はユーザと同様に使いやすさを 4 段階で評価した。下記の評価者 3 名が実験に立ち会うことなく、被験者のインタラクションデータを再生しながら使いやすさを評価した。

評価者 A: ユーザビリティ評価実務者 (視線の分析経験あり)

評価者 B: 男子大学生

評価者 C: 女子大学生

実験の結果、ユーザと評価者から得られた評価の集計結果を表 1 に示す。表 1 の集計結果にはユーザによる評価の使いやすさごとに集計した結果を示す。また、ユーザによる評価と評価者による評価の関係を分析するため、評価の対応関係を表している。例えば、ユーザによる評価の結果、「1. 使いにくい」と評価されたページは 11PV である。評価者 A がこれらのページの使いやすさを評価した結果、「1. 使いにくい」が 1PV、「1. どちらかといえば使いにくい」が 7PV、「1. どちらかといえば使いにくい」が 2PV、「4. 使いやすい」が 1PV となる。

5. ユーザと評価者の評価結果の比較分析

実験で記録したユーザによる評価結果と評価者の評価結果を比較し、評価の差異について分析を行った。ユーザと評価者の間で評価結果が一致しているか、 κ 統計量を用いて分析した。分析の結果、有意水準 0.05 でどの評価者の評価結果も被験者の評価結果と異なる評価であり、ユーザと評価者では評価が一致しておらず、差異がみられることが分かった。そのため遠隔非同期ユーザビリティテストでは評価者がユ

表 1 ユーザと評価者による評価の集計結果

Table 1 Summary of Users' evaluation and Evaluators' evaluation.

	ユーザによる 評価 (PV)	評価者 A による 評価 (PV)	評価者 B による 評価 (PV)	評価者 C による 評価 (PV)
使いやすさ		1 2 3 4	1 2 3 4	1 2 3 4
1. 使いにくい	11	1 7 2 1	2 6 2 1	1 4 1 5
2. どちらかといえば 使いにくい	11	0 4 4 3	1 2 5 3	1 2 1 7
3. どちらかといえば 使いやすい	17	0 6 8 3	1 1 9 6	1 3 1 12
4. 使いやすい	12	0 3 6 3	1 2 4 5	0 1 2 9
平均値	2.588	2.765	2.882	3.333

表 2 ユーザと評価者による評価の比較分析結果

Table 2 Comparison Analysis between Users' evaluation and Evaluators' evaluation.

分析手法	評価者 A による 評価	評価者 B による 評価	評価者 C による 評価
κ 統計量 p-value	0.437	0.169	0.802
相関の検定 p-value	0.044	0.014	0.064
ピアソンの相関係数	0.283	0.341	0.261
符号検定(片側) p-value	0.308	6.62e-05	0.008

ユーザのインタラクションの振る舞いからユーザの評価と一致した評価をすることは難しいことが分かった。

次にユーザによる評価結果と評価者による評価結果の相関の有無を検定した。検定の結果、評価者 A、B の評価結果は被験者による評価結果と相関があり、それぞれ評価者 A の相関係数が 0.283、評価者 B の相関係数が 0.341 と弱い正の相関がみられた。このことから評価者はユーザと一致した評価は難しいが、同じ傾向の評価は可能であ

ることが分かった。

ユーザと評価者による評価結果に正の相関が見られたため、評価者がユーザと比較して過大評価しているのか、符号検定を行った。すでに表 1 の平均値の結果から過大評価している傾向がみられるため片側符号検定を行った。検定の結果、有意水準 0.05 で評価者 B、C はユーザと比較して過大評価していることがわかった。しかし、実務者である評価者 A は過大評価していると言えないことがわかった。

これらの結果から、評価者がユーザと同様の評価をつけることは難しいが、実務者であれば被験者の評価結果と相関があり、過大評価はしていないという結果が得られた。

6. 考察

ユーザによる評価結果と評価者の評価結果を比較分析した結果から評価者間の評価の差異について考察する。表 1 のユーザと評価者による評価の集計結果から全ての評価者はユーザが「1」と評価したページを「1」、「2」と評価している傾向がある。しかし、ユーザビリティ評価の実務者である評価者 A に着目すると評価者 B、C と異なりユーザが「3」と評価をしたページも「2」と評価する事が多い。このことから遠隔非同期で Web ユーザビリティ評価を実施する場合、実務者である評価者 A の「やや使いにくい」と評価した Web ページを改善の対象とするといった対応が必要であると考えられる。このとき「やや使いにくい」ページまでを改善の対象に入れた場合、改善の対象となる Web ページを全体の約 2/5 の絞り込むことが可能である。

Web ユーザビリティ評価ではユーザにとって使いにくいページを評価者が見落とさないことが重要であるため、ユーザが使いにくいと評価した 11PV において評価者が問題点を指摘可能であるのか、アンケート結果から考察する。ユーザが使いにくいと評価した Web ページにおいて、使いにくいと評価した要因についてアンケートを実施した。また評価者には、「使いにくい」「どちらかといえば使いにくい」と評価した Web ページにおいて同様に要因をについてアンケートを実施した。表 3 にユーザが使いにくいと評価した 11PV における評価者間の評価の差異を示す。ユーザビリティ評価の実務者である評価者 A に着目すると、評価者 A が指摘した問題点はユーザが直面した問題につながる事がわかる。評価者による評価が高いため、問題点を指摘していない Web ページを除いてユーザが直面する問題点が指摘可能である。しかし、実務者でない評価者 B、C の場合には指摘した問題点がユーザの直面した問題につながらない場合が見られた。これらの結果からユーザビリティ評価の実務者であれば遠隔非同期 Web ユーザビリティ評価の環境であっても Web ページの評価、問題点の抽出が可能であると考えられる。しかし、そのためには対面同期での視線の分析経験や実際のユーザビリティ評価の経験が必要であると考えられる。

表 3 ユーザにとって使いにくい 11PV における評価の差異

Table 3 Differences in 11 Low Usability Pagaviews.

No.	ユーザの直面した問題	評価者 A		評価者 B		評価者 C	
		評価	指摘	評価	指摘	評価	指摘
1	募集要項はこのページにあると思った メニューの文字が小さく薄い	2	○	3	○	2	○
2	情報量（リンク）が多い	4	—	4	—	4	—
3	1 ページ以内にほしい情報がなく、スクロールしなければならない	2	○	2	○	1	○
4	このページに目的の情報があると思ったのになかった	2	○	1	○	3	×
5	採用情報と募集要項で迷った 2 回目で募集要項だとわかった	2	○	1	○	4	—
6	どのリンクが何の情報を含んでいるかわかりにくい	2	○	2	○	2	○
7	商品がメインのためほしい情報のリンクがさみしい	3	—	3	—	4	—
8	すぐに次の情報に行きつかない 画像が大きくリンクに見えない	3	—	2	×	4	—
9	下まで行ったとき他のメニューに行きづらい このページにあると思った	2	○	2	×	4	×
10	情報量が多すぎて見つけにくい リンクが多すぎる	2	○	2	○	2	○
11	このページにあると思ったけど見つからない 文字の量が多かった	1	○	2	○	2	×

評価（評価者の評価）

- 1： 使いにくい
- 2： どちらかといえば使いにくい
- 3： どちらかといえば使いやすい
- 4： 使いやすい

指摘（評価者がユーザの直面した問題を指摘可能である）

- ： 指摘した問題点の全てまたはその一部がユーザの直面した問題につながる
- ×
- ×： 指摘した問題点がユーザの直面した問題につながらない
- ： 評価が高いため、問題点を指摘していない。

7. まとめ

本研究では遠隔非同期 Web ユーザビリティ評価に向けて、ユーザの主観評価とユーザのインタラクションデータのみを用いた評価者の評価の差異について分析した。評価者はユーザとのやりとりが一切無い状態で評価した。ユーザの評価結果と評価者の評価結果を比較、分析した結果、評価者がユーザと同様の評価をつけることは難しいことがわかった。しかし、ユーザビリティ評価の専門家であれば被験者の評価結果と相関があり、過大評価していないことがわかった。またユーザビリティ評価の実務者であればユーザの直面した問題は使いやすさの遠隔非同期 Web ユーザビリティ評価の環境であっても Web ページの評価、問題点の抽出が可能であると考えられる。

今後の課題として、被験者・評価者を増やした場合の検証、具体的に指摘された問題点の差異を調査する必要がある。また被験者の視線などのインタラクションを分析する際の評価者の評価インタラクションの分析、評価者が評価しやすい被験者のインタラクション再生手法の検討が必要である。

謝辞 本研究の一部は南山大学 2011 年度パッへ研究奨励金 I-A-2 の助成を受けた。

参考文献

- 1) Web サイト価値ランキング 2010
<http://japanbrand.jp/ranking/we-ranking/we2010.html>
- 2) 大石芳裕編著: 日本企業の国際化 グローバル・マーケティングへの道, 文眞堂 (2009).
- 3) Jakob, N.: Designing Web Usability, Peachpit Press (1999).
- 4) Jakob, N. and Kara, P.: Eyetracking Web Usability, New Riders Press (2009).
- 5) Javier, S. A., Henrik, S. J., P. H. and Dan, W. H.: Low-cost gaze interaction: ready to deliver the promises, CHI EA '09 Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, ACM, pp. 4453-4458 (2009).
- 6) 中道上, 木浦幹雄, 山田俊哉, 上野秀剛: Web インタラクションの協調的可視化ツールの提案, ヒューマンインタフェースシンポジウム 2010 論文集(DVD-ROM), pp.341-344 (2010).
- 7) Ernesto, A., Ted, S., Willy, W.: Work-in-progress: Usability tool for analysis of web designs using mouse tracks, CHI '06 extended abstracts on Human factors in computing systems, pp. 484-489 (2006).
- 8) Sheree, J., Michael, E., H.: Visual attention to repeated internet images: testing the scanpath theory on the world wide web, An Eye Tracking Study. Proc. ETRA, pp. 43-49 (2002).
- 9) 木浦幹雄, 大平雅雄, 上野秀剛, 松本健一: Webjig : ユーザ行動とユーザ画面の関連付けによる動的 Web サイト利用者の行動可視化システムの開発および評価, 情報処理学会論文誌, Vol.51, No.1, pp.204-215 (2010).