

モバイルサイトにおける Web ユーザビリティ評価システムの開発

廣瀬泰弘[†] 赤池英夫[‡] 角田博保[‡]

[†]電気通信大学大学院 電気通信学研究科 情報工学専攻 [‡]電気通信大学電気通信学部 情報工学科

本研究では携帯電話での利用を想定した Web サイトのユーザビリティを自動評価するシステムの開発を行った。

現在、無数の Web サイトが携帯電話を介して様々なユーザにより様々な場面で利用されるようになった。そのような状況においては、PC に対する場合とは異なる視点から Web サイトのユーザビリティを考慮する必要がある。そこで W3C の Mobile Web Best Practice をもとに対象 Web サイトの項目スコアを算出し提示するシステムを作成した。

そしてシステムの評価のために被験者実験によって数種類の Web サイトの評価を実施した。被験者はその主観総合のスコアにより 4 つのグループに分類し、そのグループごとにシステムによる予測総合スコアと被験者による主観総合スコアの比較を行った。

Development of Mobile Web Usability Scoring System

Yasuhiro HIROSE[†] Hideo AKAIKE[‡] Hiroyasu KAKUDA[‡]

[†]Department of Computer Science, Graduate school of Electro-Communications

[‡]Department of Computer Science, The University of Electro-Communications

In this research, we developed a system assessing websites accessed mainly via cellular phones. Many people access websites by their mobile device at various scenes recently, and the way it is used is different from that of PC. In other words, we should think about the usability of websites for mobile device from a different point of view from PC. So, we developed a website scoring system based on the Mobile Web Best Practices (MWBP) from W3C and some user preferences. In an evaluation experiment, 10 subjects were classified into 4 groups by mean and variance of their subjective score, and we compared the expected score from the system with the subjective score on group-by-group basis.

1 はじめに

本研究では携帯電話での利用を想定した Web サイトのユーザビリティを W3C の Mobile Web Best Practices 1.0[1] のチェックリストを基に自動評価し、ユーザへスコアを提示するシステムを開発し、有効性について検討した。

現在、無数の Web サイトが携帯電話を介して利用されるようになった。そのような状況においては PC の場合とは異なる視点でユーザ

ビリティを考慮する必要があり、Web サイトの開発や運営には多くの時間と労力が費やされている。このような問題の解決を容易にし、ユーザに負担を掛けさせない Web サイトの開発を補助するものとして、ユーザビリティガイドラインやアクセシビリティガイドラインが存在するが、それらの多くはパソコン等のブラウザでの利用を想定した Web サイトに向けたものである。そこで本研究では携帯機器上での Web 利用を容易にする新たな標準として

W3C から公開されている Mobile Web Best Practices 1.0 (以下, MWBP) に着目し, この 60 の実践リストから 9 項目を評価対象として選択した. その評価項目を基にユーザビリティを自動評価しユーザへその予測総合スコアを提示するシステムの開発を行った.

予測総合スコアの有効性については 2 つの実験を実施することで評価を行った. まず, 1 つ目の実験では各評価項目のスコアの重みを求める事を目的とし, 各被験者に Web サイトの主観的評価を行ってもらった. 次に 2 つ目の実験では求めた重みを基に実際に予測総合スコアを算出し, そのスコアと被験者による主観総合スコアとの比較を行った.

2 関連研究

ユーザビリティを評価するシステムとして, W3C mobileOK Checker[3] がある. これは W3C mobileOK Basic Tests 1.0 に基づきタグなどを検証し, そのページがモバイル向け Web ページとして適当かどうかを検証するものである. Vigo ら [4] は, mobileOK Basic Tests に基づいた特定のデバイスにおけるモバイルウェブアクセシビリティ評価ツールを開発している. これらのシステムは特定の Web ページの記述を走査しそのページがモバイル環境に適したものかを検証するツールであり, また, ユーザビリティやアクセシビリティを評価したスコアなどの指標を示すものではない. 本システムは特定の Web ページだけではなくその Web サイト全体を総合的に検証し, そのユーザビリティやアクセシビリティを評価したスコアを提示する.

また, Sauro ら [5] は “効率性”, “有効性”, “満足度” を 4 つの変数によって, ユーザビリティを定量的に評価しそれらを 1 つのスコアへとまとめる研究を行った. 1 つの統合されたスコアを用いて表現することでユーザビリティという概念を定量的な指標としてシステムなどの評価に用いることができるが, この手法は主成分分析により各指標の重みを求めるた

め非常に多くの被験者によるデータが必要となり, また, これにより求められた指標はユーザの主観的評価を予測するものとはなっていない.

3 提案システム

3.1 システムの概要

本システムは, 携帯電話での利用を想定した Web サイトを自動的に評価するシステムである. まずシステムへの入力として解析対象 Web サイトのトップページの URL を与える. システムはこのトップページから解析する範囲内全てのリンク構造を解析する. その解析結果を基にして各評価項目ごとに Web ページを評価し各項目ごとのスコアを算出する. こうして求められた計 9 項目のスコア (項目スコア) はスコア関数によって重みを与えられ 1 つのスコア (予測総合スコア) へと統合される. これらシステムの構成を, 図 1 に示す.

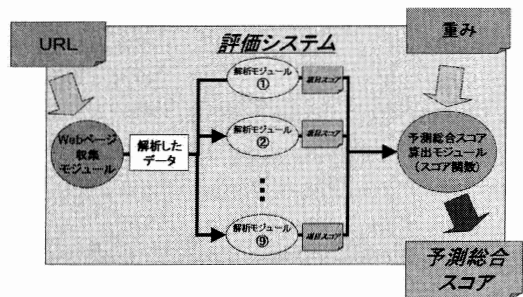


図 1: システムの構成

なお, Web サイトにおいて特定のタスクを実行する際の評価を行う場合には, トップページの入力と, タスクのゴールページの入力により解析を行う. この場合はトップページからゴールページまでの経路は, ボタンの入力数が最小になる経路を設定した.

また, 図 2 はシステムの起動時の状態, 図 3 は解析結果表示エリアを拡大したものを示したものである.

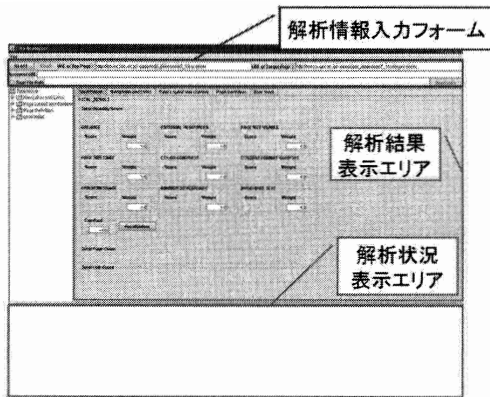


図 2: インタフェースの概観

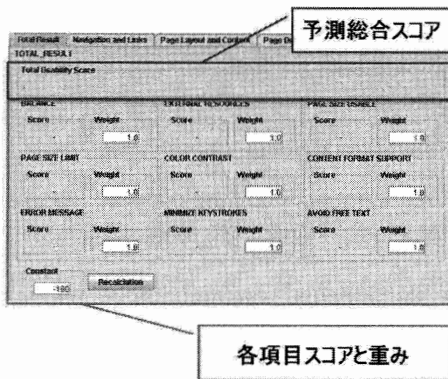


図 3: 解析結果表示エリア

3.1.1 予測総合スコア

予測総合スコアは以下の関数（以下、スコア関数）として定義する。

$$f_{score}(x_1, x_2, \dots, x_n) = \sum_{i=1}^n c_i x_i + c_0 \quad (1)$$

ただし、 $x_i (i = 1, \dots, n)$ は i 番目の項目の評価スコア、 $c_i (i = 1, \dots, n)$ は評価スコアの重み、 c_0 は定数とする。また提示するスコアの範囲は 0~100 のため、求めた値が負であるときには 0, 100 を上回るときには 100 とする。

3.2 本システムの評価対象について

この節では本研究で評価対象とした 9 項目において、それぞれの概要と評価対象の要素

について述べる。尚、□内の数字は MWBP における番号である。

項目 1. バランス [7]

携帯電話はディスプレイサイズに制限があるため、表示させている Web ページの中に大量のリンクが存在すると、ユーザに多くのスクロールを強いたり、探索に多くの時間とコストがかかってしまう。解析対象のページにおけるリンク数を基に以下の式を定義しスコア (S) の算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n t_i / l_{max} \quad (2)$$

$t_i : \max(l_{max} - l_i, 0)$,

l_i : ページ i 中に存在するリンクの数,

l_{max} : 任意で設定した 1 ページにおけるリンクの最大数

項目 2. 外部リソース [16]

画像などそのページにリンクされた外部リソースは、ネットワーク間で個々のリクエストをサーバへ要求するため、特に携帯電話など通信の大域幅の狭いデバイス環境ではページの読み込みに時間がかかってしまう。本システムでは 1 ページあたりで利用される画像リソースの数を基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n t_i / e_{max} \quad (3)$$

$t_i : \max(e_{max} - e_i, 0)$,

e_i : ページ i 中の画像リソースの数,

e_{max} : 任意で設定した 1 ページの画像リソースの最大数

項目 3. 使えるページ・サイズ [20]

1 つのページに大量の情報が掲載されページのサイズが非常に大きくなると、読み込みに過度に時間がかかるだけでなく、一般的にデバイスがその容量の制限を持っている場合が多い。本システムでは利用を想定する携帯デバイスで扱える記述（マークアップ）の最大容量と、解析対象

のページの容量を基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n t_i / m_{max} \quad (4)$$

t_i : $\max(m_{max} - m_i, 0)$,

m_i : ページ i の記述量¹,

m_{max} : 1 ページあたりの記述量の最大サイズ

項目 4. ページ・サイズの制限 [21]

携帯電話などのデバイスではメモリ容量に厳しい制限があり、ブラウザによって一度に読み込むことのできるページの容量が制限されている場合が多いため、ページ内の外部リソースの容量の合計がそのブラウザの制限を超過してしまい、必要な情報が読み込まれないという可能性もある。本システムでは利用を想定する携帯デバイスで1度に読み込むことのできる最大容量と、解析対象の画像リソースを含むページの容量を基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n t_i / p_{max} \quad (5)$$

t_i : $\max(p_{max} - p_i, 0)$,

p_i : ページ i のデバイスに読み込まれるサイズの合計,

p_{max} : 1 ページあたりのデバイスに読み込める最大サイズ

項目 5. 色のコントラスト [27]

近年の携帯電話のスクリーンはパソコンのディスプレイと比較すると良いコントラストでない場合がある。携帯電話は理想的でない照明条件下で利用されることが多いため、良いコントラストで情報を提供する必要がある。本システムでは WCAG2.0² で提示しているコントラスト比（輝度比）の達成基準を最低とし、そのページのリンクを含むテキストコンテンツの色（前景色）と背景色の輝度比を

基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \frac{t_j}{21 - r_{min}} \right) \quad (6)$$

t_j : $\max(r_i - r_{min}, 0)$,

r_{min} : コントラスト比の達成基準 (WCAG2.0 では LevelAA で 4.5),

r_j : テキスト要素 i の色とその背景色とのコントラスト比の平均

項目 6. コンテンツ・フォーマットのサポート [46]

携帯電話においては、利用可能なコンテンツのフォーマットは厳しく制限されている場合が多いため、利用できないコンテンツをサーバから転送することは、時間やお金などのコストを無駄にすることになる。本システムでは解析対象のページ中に存在するリンクのうち、想定している携帯デバイスで利用できない形式のものがどの程度あるかについて解析し、その数を基に以下の式を定義しスコア (S) の算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n c_i / l_i \quad (7)$$

c_i : ページ i におけるデバイスで利用可能（閲覧可能）なリンクの数,

l_i : ページ i 中に存在するリンクの数

項目 7. エラー・メッセージ [50]

モバイル環境に限らずユーザは求めるコンテンツや情報へ失敗無くアクセスできるとは限らないため、エラー状態から戻すナビゲーションは有効な手段であり、そのようなナビゲーション構造にすべきである。本システムでは解析対象のページまでたどったとき、その途中のページまでダイレクトに戻るることのできるリンクが、そのパスの数に対しどの程度あるかを解析し、その数を基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n b_i / p_i \quad (8)$$

¹HTML の記述容量。

²Web Content Accessibility Guidelines 2.0[2]

b_i : ページ i までたどってきたページへ戻ることのできるリンクの数,

p_i : たどってきたページの数

なお、基準となるパスは解析対象のページまでの最短経路を利用した。

項目 8. キー入力の数最小化 [54]

ポインティングデバイスを持たない携帯電話のようなモバイル環境では全てのブラウザや Web 上のコンテンツの操作を方向キーを含むキーで操作するため、キーストローク数は必然的に多くなってしまふ³。本システムでは解析対象のページまでたどるのに必要なボタンの押下回数を基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n t_i / k_{max} \quad (9)$$

t_i : $\max(k_{max} - k_i, 0)$,

k_i : トップページからページ i までのキー入力数,
 k_{max} : 任意で設定したキー入力数の最大 (基準) 数

なお、基準となるパスは解析対象のページまでの最短経路を利用した。

項目 9. フリー・テキストを避ける [55]

モバイルデバイスの文字入力には入力文字や入力方法の切り替えの必要があったり、ポインティングデバイスを持たないために全てをキー操作で行う必要があるなど入力の制限があるため、Web サイトに配置されているユーザインタフェースでは文字入力などを最小限にすべきである。本システムでは 1 ページ中のテキスト入力フォームの最大値を任意で設定し、解析対象のページ中に存在するテキスト入力フォームの数を基に以下の式を定義しスコアの算出を行った。

$$S = \frac{100}{n} \sum_{i=1}^n t_i / f_{max} \quad (10)$$

³マウスを用いた場合のキーストロークとはマウスのクリック数とキーボードによるショートカット操作の回数に相当すると考える。

t_i : $\max(f_{max} - f_i, 0)$,

f_i : ページ i 中に存在するテキスト入力フォームの数,

f_{max} : 任意で設定した 1 ページにおけるテキスト入力フォームの最大数

4 調査実験

予測総合スコアを算出するための各項目スコアの重みを求めるために被験者実験を実施した。

本実験のために 9 つの評価項目につき 3 段階でパラメータを変化させた計 27 種類の Web サイトを作成し、10 名の被験者によって、それぞれの Web サイトの 0~100 点で主観総合スコアを得た。

4.1 実験環境

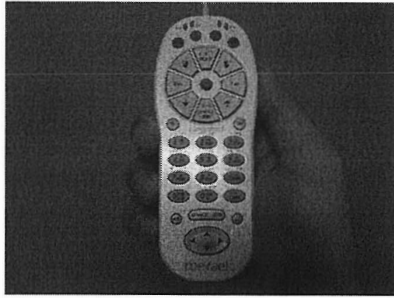
本実験で使用した実験環境を挙げる。本実験では Web サイトの表示に、PC で実行可能な携帯電話のブラウザとして、i モード HTML シミュレータ 2 (図 4(a)) を使用した。これは、通信速度やメモリ容量の制限などの設定をすることができ、実際の通信速度や読み込み時間のシミュレーションを行うことができる。また、入力には株式会社メヴァエル製の Keiboard⁴ (図 4(b)) を使用した。

- ノート PC (DELL INSPIRON 700 m)
- i モード HTML シミュレータ 2
Version 7.0
- Keiboard PA-101

4.2 被験者

被験者は男性 8 名、女性 2 名の計 10 名で、全員本学の情報工学科および情報工学専攻の学生である。また、全員が携帯電話を所持し、携帯電話によるウェブサイトの閲覧経験がある。

⁴2009 年 1 月現在 Keiboard (PA-101) は販売されていない。



(a)



(b)

図 4: Keyboard と i モード HTML シミュレータ 2

4.3 被験者の分類

表 1 は使用した 27 種類の Web サイトでの被験者による主観総合スコアをまとめたものである。

この結果に基づき、10 名の被験者を平均と標準偏差によって、以下の 4 つのグループに分類した。

グループ A [低めの点数で、レンジが狭い]

- 平均：60 未満
- 標準偏差：20 未満
- 被験者：2, 7

グループ B [低めの点数で、レンジが広い]

- 平均：60 未満
- 標準偏差：20 以上
- 被験者：10

グループ C [高めの点数で、レンジが狭い]

表 1: 主観総合スコアの平均と標準偏差

被験者	1	2	3	4	5
平均	68.78	54.07	62.30	71.67	68.15
標準偏差	23.06	18.19	29.55	24.02	25.43
被験者	6	7	8	9	10
平均	70.93	55.19	62.74	62.85	59.11
標準偏差	19.17	17.62	28.76	15.73	23.13

- 平均：60 以上
- 標準偏差：20 未満
- 被験者：6, 9

グループ D [高めの点数で、レンジが広い]

- 平均：60 以上
- 標準偏差：20 以上
- 被験者：1, 3, 4, 5, 8

4.4 算出方法

まず 27 種類のサイトにおいてシステムにより 9 項目それぞれの項目スコアを算出した。重みの算出は各グループ毎に行うため、27 のスコアをグループ内で平均し、この 27×4 グループの主観総合スコアとシステムによる 27×9 の項目スコアを使って、以下に示した式によって各項目スコアの重みを算出した。

システムの算出した予測総合スコア（スコア関数）を、被験者による主観総合スコアに対して最小二乗法により近似すると以下の J が成り立つ。

$$J = \sum_j^m (S_j - (\sum_i^n c_i x_i + c_0))^2 \quad (c_0 \text{ は定数})$$

よって J を最小化する $c_k (0 \leq k \leq n)$ を求めるため J を c_k についてそれぞれ偏微分すると以下の 2 式が求められる。

$$\left(\sum_j^m x_{kj}\right)c_0 + \sum_i^n \left(\sum_j^m x_{kj}x_{ij}\right)c_i = \sum_j^m x_{kj}S_j \quad (11)$$

$$m c_0 + \sum_i^n \left(\sum_j^m x_{ij}\right)c_i = \sum_j^m S_j \quad (12)$$

となる。以上、式 11 と式 12 から得られる連立一次方程式を解くことで c_k が求まる。ただし、今回は $n = 9, m = 27$ である。

5 評価実験

各被験者グループごとに求めた各項目スコアの重みを利用して、スコア関数により各被験者ごとに計算された予測総合スコアとその被験者による Web サイトの主観総合スコアを比較した。本実験のために 3 種類の Web サイトを作成し、重みを求めたときと同じ 10 名の被験者によって、それぞれの Web サイトの 0 ~ 100 点での主観総合スコアを得た。なお、実験環境及び被験者は調査実験と同じであるが、1 週間後に行った。

5.1 スコアの算出に用いた項目とその重み

予測総合スコアを求める重みは、9 項目の全ての組み合わせ (511 通り) から、その重みを利用して算出された予測総合スコアの予測誤差 (RMS⁵) が主観総合スコアと最も近くなる組み合わせを求めた。以下に、各グループ毎の予測総合スコア算出に用いた項目と、その重みを示した (表 2)。

グループ A

項目 : 1, 3, 5, 6, 9

グループ B

項目 : 1, 2, 5, 6, 8

グループ C

項目 : 1, 2, 4, 5, 6, 8

グループ D

項目 : 1, 2, 5, 6, 8

5.1.1 実験結果

被験者グループごとに求めた重みを利用して予測総合スコアを算出した。なお、3 つのサイトでの主観総合スコアにより 10 人の被験者を各グループへ 4.3 節と同様にして再配分を行った上で、スコアの比較を行った。その結果は以下のようになった (表 3, 4, 5)。

グループ A 被験者 : 3, 7

グループ B 被験者 : 1, 2, 5

グループ C 被験者 : 4, 6, 8, 9, 10

グループ D 被験者 : 該当者なし

⁵二乗平均平方根 (Root Mean Square)

表 2: 各グループにおけるスコアの重み

c_0	-58.49	c_0	-163.43
c_1	0.50	c_1	1.21
c_3	-0.07	c_2	0.58
c_5	0.34	c_5	0.58
c_6	0.48	c_6	0.87
c_9	0.30	c_8	-0.52
グループ A		グループ B	
c_0	-71.46	c_0	-177.67
c_1	0.48	c_1	0.80
c_2	0.29	c_2	1.00
c_4	0.07	c_5	0.49
c_5	0.16	c_6	0.859
c_6	0.78	c_8	-0.23
c_8	-0.14	グループ D	
グループ C			

表 3: グループ A におけるスコアとその RMS

サイト	予測総合スコア	被験者 3	被験者 7
1	70.9	60	70
2	41.9	40	50
3	62.3	40	60
RMS	-	14.35	4.88

6 考察

6.1 間隔尺度としてのスコア

システムの算出した予測総合スコアを量的な尺度として適当かを検証するために、被験者ごとにピアソンの積率相関係数を求めた。結果を表 6 を示す。これから読み取れるように被験者 1 と 9 以外で強い正の相関関係が見られた。以上より、間隔尺度としては主観的な評価に対し予測ができていると考えられる。しかし、結果からも分かるが、予測値としての誤差がまだあるため、データの数を増やし検証を進めていきたい。

表 4: グループ B におけるスコアとその RMS

サイト	予測総合スコア	被験者 1	被験者 2	被験者 5
1	77.9	80	80	75
2	33.2	40	30	35
3	75.6	40	60	65
RMS	-	20.95	9.25	6.43

表 5: グループ C におけるスコアと RMS

サイト	予測総合スコア	被験者 4	被験者 6	被験者 8
1	76.0	90	90	80
2	62.7	70	55	65
3	70.6	80	85	70
RMS	-	10.62	12.41	2.71

サイト	予測総合スコア	被験者 9	被験者 10
1	76.0	60	75
2	62.7	75	40
3	70.6	55	70
RMS	-	14.74	13.11

表 6: 主観総合スコアと予測総合スコアの相関係数 r

被験者	1	2	3	4	5
r	0.54	0.94	0.73	0.99	0.98

被験者	6	7	8	9	10
r	0.96	0.97	0.95	-0.79	0.96

6.2 順序尺度としてのスコア

次に、システムの算出した予測総合スコアが順序尺度として適当かを検証するために、被験者ごとにスピアマンの順位相関係数を求めた。結果を表 7 を示す。これから読み取れるよう

表 7: 主観総合スコアと予測総合スコアの順位相関係数 r_s

被験者	1	2	3	4	5
r_s	1.00	1.00	1.00	1.00	1.00

被験者	6	7	8	9	10
r_s	1.00	1.00	1.00	-0.50	1.00

に強い順位相関が見られた。以上より、順序尺度として予測総合スコアを捉えたとき非常に良い精度で予測できている考えられる。

7 まとめ

本研究では携帯電話での利用を想定した Web サイトを解析しそのユーザビリティを自動評価し、スコアをユーザに提示するシステムの開発を行った。また、予測総合スコアやスコア関数の妥当性を検証するために評価実験を行った。

解析の結果、主観総合スコアとシステムの算

出した予測総合スコアの間に数量的な相関が見られ、さらに順序相関も強いということが分かったが、主観的評価との数量的誤差が大きくなってしまったといった問題があった。これらの結果を踏まえ今後の課題として、様々なバリエーションを持つウェブサイトによる被験者実験の実施を行い、ユーザのデータを収集していきたい。加えて、携帯電話などのデバイスにおける Web ブラウジングにおいて、ある要素がどの程度存在するとユーザのパフォーマンスやユーザエクスペリエンスに影響を及ぼすのかといった調査を行い、評価項目の更なる検討やそれらの定量的な評価方法の提案を行っていく必要がある。

参考文献

- [1] Mobile Web Best Practices 1.0 Basic Guidelines, W3C Recommendation, 2008, <http://www.w3.org/TR/mobile-bp/>
- [2] Web Content Accessibility Guidelines 2.0, W3C Recommendation, 2008, <http://www.w3.org/TR/WCAG20/>
- [3] W3C mobileOK Checker [Beta], <http://validator.w3.org/mobile/>
- [4] Markel Vigo, Amaia Aizpurua, Myriam Arrue, "Evaluating web accessibility for specific mobile devices", Proceedings of W4A2008, 2008, pp.65-72
- [5] Jeff Sauro, Erika Kindlund, "A Method to Standardize Usability Metrics Into a Single Score", Proceedings of CHI 2005, 2005, pp.401-409