

学術用語解説ウェブページの良否評定のための 分かり易さ・見易さ因子の分析

Analyzing Factors of Beginner Friendliness / Visual Intelligibility
for Overall Measurement of Web Pages Explaining Academic Concepts

廣花 智遥 *¹
Chiharu Hirohana

岡田 心太郎 *¹
Shintaro Okada

宇津呂 武仁 *²
Takehito Utsuro

河田 容英 *³
Yasuhide Kawada

神門 典子 *⁴
Noriko Kando

*¹筑波大学大学院システム情報工学研究科
Grad. Sch. Sys. & Inf. Eng. Univ. of Tsukuba

*²筑波大学システム情報系
Fclty. Eng. Inf. & Sys. Univ. of Tsukuba

*³(株) ログワークス
Logworks Co., Ltd.

*⁴国立情報学研究所
National Institute of Informatics

When we beginners study academic concepts on the Internet, the beginner friendly Web pages are often not shown by the search engine. In order to reduce the burden due to that, it is necessary to quantitatively evaluate “beginner friendliness” as well as “visual intelligibility” in Web pages explaining academic concepts. In this paper, we manually analyze and explore the cause of the lack of “beginner friendliness” as well as “visual intelligibility”. Especially, we focus on the cases where the Web page lacks either “beginner friendliness” or “visual intelligibility”, even though the Web page satisfies other one. Then, we manually analyze factors of the lack of “beginner friendliness”/“visual intelligibility” and report their statistics in detail.

1. はじめに

インターネット、さらにはパソコンやスマートフォンが普及した現在、ウェブ上には数多くのコンテンツが存在する。平成29年度版総務省情報通信白書^{*1}によれば、各世帯におけるパソコンとスマートフォンの保有率は同程度であり、特に13~19歳の8割以上がスマートフォンを所持している。この年代は中学生から大学生に該当しており、学習の手段としてインターネットを頻繁に利用している。学術用語を学ぶためのコンテンツも数多く存在するため、検索エンジンに学術用語を入力するだけで関連ページを見つけることができる。しかし、検索上位に「分かりやすい」ウェブページが表示されるとは限らず、一件ずつ閲覧し複数のウェブページを見比べる非効率な作業が必要となる。また、「見づらい」ウェブページも存在するため、比較作業を途中で諦める可能性も高い。

この現状に対する取り組みとして、[塩川 19]では、ウェブ閲覧者にとって見易く、分かり易く、かつ、役に立つページであるために充足すべき因子を深層学習によって機械的に特定し、「解説型ウェブページ」の分かり易さと見易さを自動評定する仕組みを提案している。ただし、「機械的に特定」するため、その基準が人間による直観にどの程度一致しているかを慎重に吟味することが必要となる。そこで本論文では、[塩川 19]において深層学習のための教師用事例として蓄積された学術用語解説ウェブページを対象として、人手によって評定された学術用語解説ウェブページの「分かり易さ」・「見易さ」の詳細な因子についての分析を行う。本論文では、特に、「分かり易さ」と「見易さ」のうち、「分かり易さ」のみが充足され、「見易さ」が充足されない場合、および、逆に、「見易さ」のみが充足され、「分かり易さ」が充足されない場合に焦点を当て、それぞれ「見易さ」あるいは「分かり易さ」を損なう因子群を網羅的に分析した結果を報告する。

2. 参照学術用語ウェブページ集合の作成

2.1 対象学術分野および用語

本論文では、文章の分かり易さ、レイアウトの見易さ、全体評定の三つにおける評定基準が類似していると考えられる理工学学術分野を対象とする。具体的には、統計、線形代数、物理、化学、IT、プログラミング、生物の7分野を対象とする。また、各分野における学術用語の内、主に高校3年生または大学レベルの学術用語15語(表2)を評価対象とする。

2.2 データセット作成手順

前節で対象とする各学術用語を検索クエリとし、ウェブ検索の検索結果上位10件の用語解説ウェブページを収集する。その際、検索結果リンク先のウェブページにアクセスできない場合、検索結果がpdfファイルである場合、および、複数の学術分野に横断して検索される計5サイト(Wikipedia^{*2}、コトバンク^{*3}、Weblio^{*4}、Yahoo!知恵袋^{*5}、goo辞書^{*6})は除外対象とする。除外された分については、検索順位10位より下位のページから順に補充する。以上のウェブページ収集結果に対して、次節の評定基準に基づき、分かり易さ、見易さ、および、全体評定の各評定を行った結果を付与する。

2.3 用語解説ウェブページの人手評定基準

本節では、各々の用語解説ウェブページに対して、人手で分かり易さ、見易さ、および、全体の評定を行う際の評定基準(表1)について述べる。分かり易さ、および、見易さの評定基準においては、絶対条件と優先付け項目の二つを区別して設け、絶対条件を満たした上で複数の優先付け項目を満たす場合に、当該評定を満たすと評定する。

さらに、分かり易さ、および、見易さ両方の評定結果に基づき、ウェブページ全体の評定を行う。ここで、特に、分かり易さ、あるいは、見易さのいずれかの評定のみが充足され、もう一方が充足されない場合でも、充足されない方において、

連絡先: 廣花 智遥, 筑波大学大学院システム情報工学研究科,
〒305-8573 茨城県つくば市天王台1-1-1, 029-853-5427

*¹ <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc111110.html>

*² <https://ja.wikipedia.org/>

*³ <https://kotobank.jp/>

*⁴ <https://ejje.weblio.jp/>

*⁵ <https://chiebukuro.yahoo.co.jp/>

*⁶ <https://dictionary.goo.ne.jp>

表 1: 用語解説ウェブページの分かり易さ・見易さの評価基準

(a) 分かり易さの基準

絶対条件	a	用語の定義がある	対象の学術用語に関する定義がウェブページ内に含まれている。
優先付け項目	b	文章の文字数・単語数・専門用語数	ウェブページ内に含まれる文章の、文字数・単語数・専門用語数が極端に多すぎず少なすぎない。
	c	文の長さ	ウェブページ内に含まれる1文の長さが、極端に長すぎず短すぎない。
	d	例がある	対象の学術用語に関する例題がウェブページ内に含まれている。
	e	関連用語がある	対象の学術用語に関連する用語がウェブページ内に含まれている。

(b) 見易さの基準

絶対条件	f	文字と図の比率	表や図に対し、文字が極端に多すぎず少なすぎない。
優先付け項目	g	色	黒色や原色など、文章を読む際に読みづらいと感じるような背景色を使用していない。
	h	コントラスト	白色の背景に黄色の文字等、文章を読む際に読みづらいと感じるような配色をしていない。
	i	図がある	用語解説や例示のために使用されている図が1つ以上含まれている。
	j	代表的な数式がある	一般的に数式と呼べるものを含んでいる。アルファベット1文字などは対象としない。
	k	文章と数式の比率	文に対して極端に数式が多くないかどうか。
	l	字の大きさ	文章を読む際に読みづらいと感じるほど、字が小さくない。

表 2: 学術分野ごとの対象学術用語および評価一覧

分野	対象とする学術用語 (全 15 語の用語)	評定						計
		全体		分かり易さ		見易さ		
		×	○	×	○	×	○	
統計	事後確率, 事前分布, 正規分布, F 分布, 信頼区間, 標準偏差, 主成分分析, 回帰分析, 自己回帰モデル, ポアソン分布, マルコフ連鎖モンテカルロ法, 確率密度関数, コーシー分布, 分散, パレート分析	54	96	48	102	56	94	150
線形代数	階数, 共役勾配, 行列式, クラメルの公式, クロネッカーのデルタ, 三角行列, 正規直交基底, 対角化, 直交行列, 特性多項式, 二次形式, ノルム, メネラウスの定理, ヤコビ行列, 内積	99	51	88	62	88	62	150
物理	電気力線, 張力, 慣性の法則, 遠心力, 電波, 電流, 万有引力, 交流, 音波, ホイートストンブリッジ, 反発係数, 相互誘導, 正電荷, 速度, 変圧器	101	49	90	60	98	52	150
化学	イオン結合, エステル, カルボン酸, ケトン, 化学反応式, 化学平衡, 共有結合, 合成高分子, 酸化還元, 遷移元素, 典型元素, 天然高分子, 燃料電池, 物質の三態, 有機化合物	92	58	85	65	87	63	150
プログラミング	C 言語, Java, エスケープシーケンス, コマンドライン引数, スコープ, フィールド値, ポインタ, メソッド, 繰り返し処理, 構造体, 算術演算子, 条件分岐, 配列変数, 文字列, 論理演算	86	64	79	71	92	58	150
IT	API, DBMS, HTML, IP アドレス, JDBC, RDB, SDK, SQL, Unicode, URL, スコープマネジメント, ステークホルダーマネジメント, タイムマネジメント, ナレッジマネジメント, リスクマネジメント	76	74	75	75	49	101	150
生物	DNA, ショウジョウバエ, 原核生物, 減数分裂, 光合成, 細胞, 葉緑体, ミトコンドリア, 遺伝子, タンパク質, マクロファージ, RNA, ヘモグロビン, ナトリウムポンプ, コケ植物	112	38	98	52	71	79	150
計	—	620	430	563	487	541	509	1,050

表 3: 「全体」・「分かり易さ」・「見易さ」の評価の組み合わせの統計

評定			分野									
全体	分かり易さ	見易さ	統計	線形代数	物理	化学	プログラミング	IT	生物	計		
×	×	×	34	57	57	55	53	33	60	349	(33.2 %)	
		○	10	29	32	30	26	40	36	203	(19.3 %)	
	○	×	10	12	12	6	6	2	7	55	(5.2 %)	
		○	0	1	0	1	1	1	9	13	(1.2 %)	
○	×	×	0	0	0	0	0	0	0	0	(0 %)	
		○	4	2	1	0	0	2	2	11	(1.0 %)	
	○	×	12	19	29	26	33	14	4	137	(13.0 %)	
		○	80	30	19	32	31	58	32	282	(26.9 %)	
計			150	150	150	150	150	150	150	1,050	(100 %)	

表 1(a), あるいは, 表 1(b) に示す優先付け項目のうちの一部を充足する場合には, 全体評価結果が「良」と評価されている事例が多く観測される。

2.4 データセット

作成されたデータセットにおける「全体評価」, 「分かり易さ」, 「見易さ」, の各々における良否評価分布を表 2 に示す。表 2 においては, 各学術分野における対象学術用語 (各 15 語) もあわせて示す。

3. 評価結果における見づらさ・分かりづらさ因子群の分析

3.1 評価結果の分析

表 2 において, 「全体評価」, 「分かり易さ」, 「見易さ」各々の良否二値評価結果の組み合わせ全 8 通りの分布を表 3 に示す。「分かり易さ」, 「見易さ」とともに「否」評価を覆し, 「全体評価」を「良」評価とする事例は 0 件であることから, 2.3 節の基準のうちの全体評価の基準が反映されていることが確認できる。「分かり易さ」, 「見易さ」とも「否」評価の場合は「全体評価」も「否」評価となり, 逆に, 「分かり易さ」, 「見易さ」とも「良」評価の場合は「全体評価」も「良」評価となる傾向にある。また, 「分かり易さ」が「否」評価, 「見易さ」が「良」評価の場合は「全体評価」は「否」評価 19.3%, 「良」評価 1.0% であり, 圧倒的に「否」評価となる傾向にある。逆に, 「分かり易さ」が「良」評価, 「見易さ」が「否」評価の場合は「全体評価」は「否」評価 5.2%, 「良」評価 13.0% と混在しており, 「分かり易さ」「良」評価の影響の方が強い傾向にある。

3.2 見づらさ因子群の分析

本節では, 表 3 中で「分かり易さ」が「良」評価, 「見易さ」が「否」評価の場合に「全体評価」が「否」評価となる 55 事例 (5.2%) に着目し, 表 4 に示す「見づらさの因子」19 個の分布を図 1 に示す。図 1 においては, 当該因子が「見づらさ」の原因となる主な項目である場合, 逆に, 「見づらさ」の原因としては補助的項目である場合を区別して示す。

この結果から, 因子「色コントラスト」および因子「項目間空行幅」の順に該当ページ数が多い。また, 因子「色コントラスト」においては, 補助的項目となる事例はなく, 全事例において「見づらさ」の原因となる主な項目となっている。一方, 「二段組」や「リンク過多」等は観測数が最下位相当である。

3.3 分かりづらさ因子群の分析

本節では, 表 3 中で「分かり易さ」が「否」評価, 「見易さ」が「良」評価の場合に「全体評価」が「否」評価となる 203 事例 (19.3%) のうちの 55 件に着目し, 表 5 に示す「分かりづらさの因子」12 個の分布を図 2 に示す。図 2 においては, 当該因子が「分かりづらさ」の原因となる主な項目である場合, 逆に, 「分かりづらさ」の原因としては補助的項目である場合を区別して示す。この結果から, 因子「用語解説なし」および因子「解説なしの関連用語多用」の順に該当ページ数が多い。

4. 関連研究

本論文に関連して, [春日 18, Han18] においては, 用語解説ウェブページの見易さ・分かり易さの判定を行うにあたって, 特に, HTML 構造に着目した素性を用いることにより, 見易さ・分かり易さを自動判定する手法を提案している。また, [岡田 19] においては, 「見易さ」自動判定のための深層学習を導入するとともに, その判定理由を可視化している。この可視化結

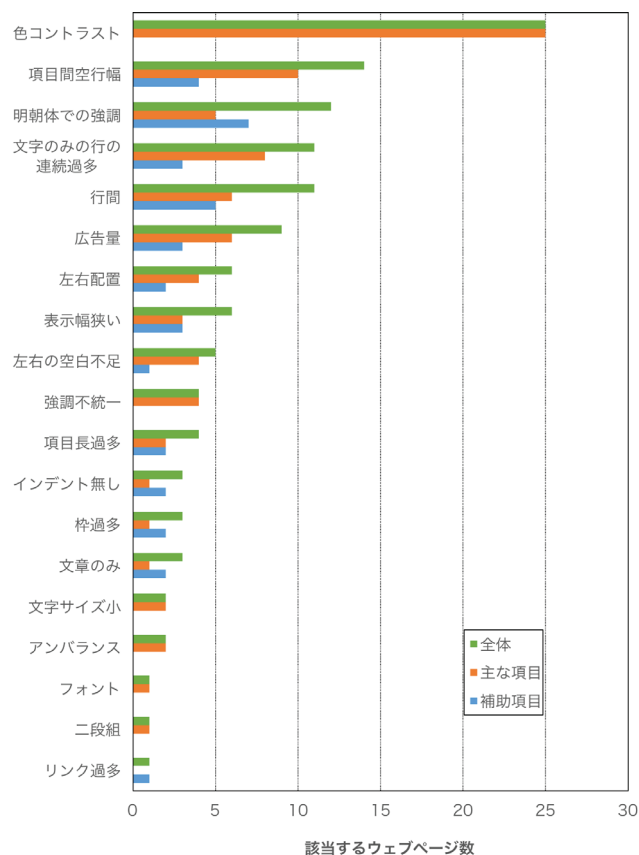


図 1: 用語解説ウェブページの見づらさの因子群の分布 (「全体評価×, 分かり易さ○, 見易さ×」である 55 例)

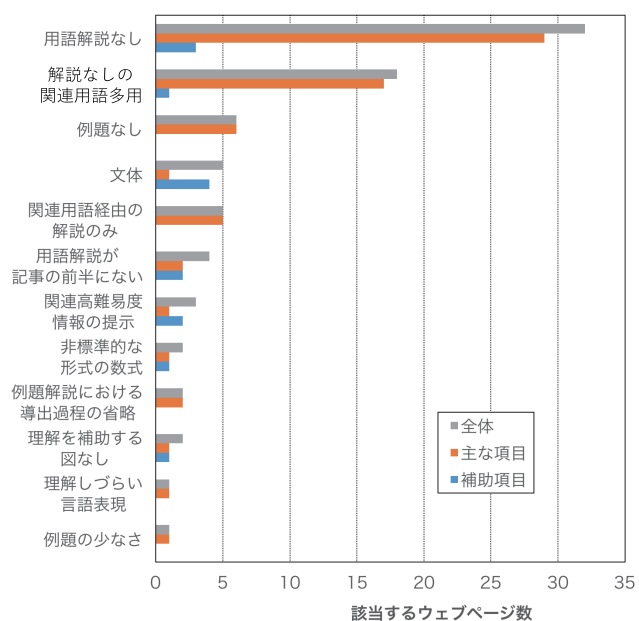


図 2: 用語解説ウェブページの分かりづらさの因子群の分布 (「全体評価×, 分かり易さ×, 見易さ○」である 55 例)

表 4: 用語解説ウェブページの見づらさの原因となる因子

	見づらさの因子	詳細
1	色コントラスト	背景色と文字色のバランスが悪い。蛍光的または刺激の強い色の枠・文字を利用。必要以上の色遣い。
2	項目間空行幅	現項目終了位置と次項目開始位置間の空行幅が不適切。
3	明朝体での強調	明朝体で強調を行なっている。
4	文字のみの行の連続過多	行数が多い、または、空行で区切られた(複数の)段落連続数が多い。
5	行間	区切り線や文章の間の行間が広い、または、狭い。
6	広告量	1 ページ中に複数の広告が存在するなど、広告が多い。
7	左右配置	通例的に、横書きは左から読む。しかし、重要な内容が右に、そうでないものが左に書かれている。
8	表示幅狭い	ブラウザウィンドウに対して、ウェブページの表示幅が狭い。
9	左右の空白不足	コンテンツがブラウザウィンドウの端に寄り過ぎてている。
10	強調不統一	強調の仕方(太字、マーカーなど)が同一ページ内に複数ある。
11	項目長過多	1 項目が長く、項目間の行数が多すぎる。
12	インデント無し	インデントが無く、どこからどこまでがどのコンテンツ(文章、図、数式、箇条書きなど)かが不明瞭。
13	枠過多	必要以上に枠を利用している。
14	文章のみ	文章以外のコンテンツ(一目で認識でき、明白に文字でないと判断できるもの)が無い。
15	文字サイズ小	文字サイズが小さい。
16	アンバランス	タイトルなど、内容として重要でないものの存在感が過度に大きい。
17	フォント	フォントの線の太さを使い分けておらず、一目では題目と内容の区別が困難。
18	二段組	二段組になっている。
19	リンク過多	多数のリンクが連続的に列挙されている。

表 5: 用語解説ウェブページの分かりづらさの原因となる因子

	分かりづらさの因子	詳細
1	用語解説なし	用語の応用先の紹介やプログラム実装のみなどに留まっており、用語解説が全くされていない。
2	解説なしの関連用語多用	専門的な関連用語や式が、何の説明もなく頻出。
3	例題なし	数値例を用いた具体的な例題の描写がない。
4	文体	文体が論文調で堅い。
5	関連用語経由の解説のみ	関連用語を介した解説のみであり、関連用語を用いない直接的解説が無い。
6	用語解説が記事の前半にない	用語解説箇所が記事の後半にある。
7	関連高難易度情報の提示	用語に関連する高難易度な内容も記述。
8	非標準的な形式の数式	数式の表現形式が非標準的な形式である。
9	例題解説における導出過程の省略	例題の解説において導出過程が省略されている。
10	理解を補助する図なし	文章による解説の理解を補助する図がない。
11	理解しづらい言語表現	解説部分に理解困難な言語表現を含む。
12	例題の少なさ	例題が少なさが原因で、理解が妨げられる。

果と本論文の分析結果を照らし合わせることによって、より重要性の高い属性を重視し、自動評価精度の改善が期待できる。一方、[小林 05, Krug16]においては、分かり易く見やすいウェブサイトの作成方針について、詳細な考え方が述べられている。

5. おわりに

本論文では、学術用語解説ウェブページに対して、「文章の分かり易さ」および「ウェブページの見易さ」の観点から人手評価を行い、データセットを作成した。本論文では、特に、「分かり易さ」と「見易さ」のうち、いずれか片方のみが充足され、もう片方が充足されない場合に焦点を当て、「見易さ」および「分かり易さ」を損なう因子群を網羅的に分析した。そして、それらの因子群のうち、「色のコントラスト」の因子が見づらさに、「用語解説なし」および「解説なしの関連用語多用」の因子が分かりづらさに、それぞれ、大きな影響を与えていることを示した。本論文の分析をふまえ、今後は表 3 における各良否組(全 8 通りのうち、事例数 0 および「全体」「分かり易さ」「見易さ」全て○を除く残り 4 通り)を対象として、「分かり易さ」、あるいは、「見易さ」を妨げる因子の分析を行う必要がある。

参考文献

- [Han18] Han, B., Shiokawa, H., Kawaguchi, K., Utsuro, T. and Kawada, Y.: Measuring Beginner Friendliness of Chinese Web Pages explaining Academic Concepts using HTML Structures, 第 32 回人工知能学会全国大会論文集 (2018).
- [春日 18] 春日孝秀, 塩川隼人, 韓炳材, 宇津呂武仁, 河田容英: HTML 構造上の特徴を利用した学術用語解説ウェブページの分かり易さの自動評価, 第 10 回 DEIM フォーラム論文集 (2018).
- [小林 05] 小林隆, 榎原直樹, 関根千佳, 遊間和子: 市民にやさしい自治体ウェブサイト: 構築から運用まで, NTT 出版株式会社 (2005).
- [Krug16] Krug, S.: 超明快 Web ユーザビリティ: ユーザに「考えさせない」デザインの法則, 株式会社ビー・エヌ・エヌ新社 (2016).
- [岡田 19] 岡田心太郎, 塩川隼人, 韓炳材, 廣花智遥, 宇津呂武仁, 河田容英: 深層学習による学術用語解説ウェブページの見易さ自動評価結果の理由提示, 第 11 回 DEIM フォーラム論文集 (2019).
- [塩川 19] 塩川隼人, 岡田心太郎, 韓炳材, 廣花智遥, 宇津呂武仁, 河田容英: 深層学習を用いた学術用語解説ウェブページの分かり易さ・見易さの自動評価, 第 11 回 DEIM フォーラム論文集 (2019).