

共起語分析に基づく業界別ウェブコンテンツ特性の考察

著者	柏崎 剛	Name	Tsuyoshi Kashiwazaki
法人	株式会社コンデンシャル	Affiliation	Consential Co., Ltd.
言語	日本語	Language	Japanese
Xアカウント	@tkashiwazaki2	X Account	@tkashiwazaki2
ORCID	0009-0000-2545-3055	ORCID	0009-0000-2545-3055
バージョン	1.0.1	Version	1.0.1
発行日	2025年4月10日	Publication	April 10, 2025
サイト	SEO対策研究室	Website	SEO Strategy Research Laboratory
メール	info@tsuyoshikashiwazaki.jp	Email	info@tsuyoshikashiwazaki.jp
設立	2012年6月26日	Founded	June 26, 2012
所在地	〒151-0053 東京都渋谷区代々木2-26-2 第二桑野ビル5D	Address	5D, Daini Kuwano Building, 2-26-2 Yoyogi, Shibuya-ku, Tokyo 151-0053, Japan
電話	03-6276-4579	Phone	+81-3-6276-4579
FAX	03-3299-1134	Fax	+81-3-3299-1134
ライセンス	CC BY 4.0	License	CC BY 4.0

要旨 : abstract

本研究では、IT、エンターテインメント、医療、金融という4つの主要業界における検索上位サイトの共起語分析を通じて、各業界のウェブコンテンツ特性を明らかにすることを目的とした。各業界を代表する検索キーワードで上位表示された10サイトの共起語リスト（各100語）について、共通基盤となる語彙の割合、共通語彙の上位ランクへの集中度、主要品詞（名詞）の構成比率、数値・記号を含む語彙の存在率という4つの指標を分析した。その結果、業界間で顕著な差異が見られ、これらの数値的特徴から各業界の情報提供の傾向や専門性、ユーザーの情報ニーズについての重要な示唆が得られた。本研究は、共起語分析がウェブコンテンツの特性理解やSEO戦略の策定、業界ごとの情報流通の特徴把握において有効な手法であることを示している。

1. 序論

1.1 研究背景

検索エンジン最適化（SEO）において、ページの内容と検索意図の適合性は極めて重要な要素である。検索エンジンは、ユーザーが入力したキーワードに関連する情報を含むページを検索結果の上位に表示する傾向があるため、ページ内の語彙の使用状況は重要な分析対象となる。その中でも、特定のキーワードと共に頻出する語彙（共起語）の分析は、コンテンツ最適化やユーザーの検索意図の理解に有効とされている。また、異なる業界では、検索上位に表示されるサイトの特性や、ユーザーが求める情報の性質が異なる可能性がある。こうした業界ごとの特性を共起語分析を通じて明らかにすることは、効果的なウェブコンテンツ戦略の構築に寄与する可能性がある。

1.2 研究目的

本研究の目的は、異なる業界における検索上位サイトの共起語分析データから、各業界および業界横断的な傾向・特色を明らかにし、共起語分析がウェブコンテンツ、SEO戦略、業界特性理解に与える有意義な示唆を提示することである。具体的には、IT、エンターテインメント、医療、金融という4つの主要業界について、特定の指標を用いた分析を行い、業界ごとの特徴や業界間の比較を通じて、ウェブコンテンツの特性や情報提供の傾向について考察する。

1.3 分析対象と手法

本研究では、IT、エンターテインメント、医療、金融の4業界において、各業界を代表する検索キーワードで上位表示された10サイトの共起語リスト（各100語）を分析対象とした。分析には以下の4つの指標を用いた：

- 共通基盤となる語彙の割合：上位10サイト間で共通して出現する語彙の割合
- 共通語彙の上位ランクへの集中度：共通語彙が検索結果の上位に集中している度合い
- 主要品詞（名詞）の構成比率：共起語リスト内における名詞の割合
- 数値・記号を含む語彙の存在率：数字や特殊記号を含む語彙の割合

これらの指標に基づいて、各業界の特徴を定量的に分析し、業界間の比較を行った。また、得られた数値データから、各業界のウェブコンテンツにおける情報提供の傾向や専門性、ユーザーの情報ニーズについての考察を行った。

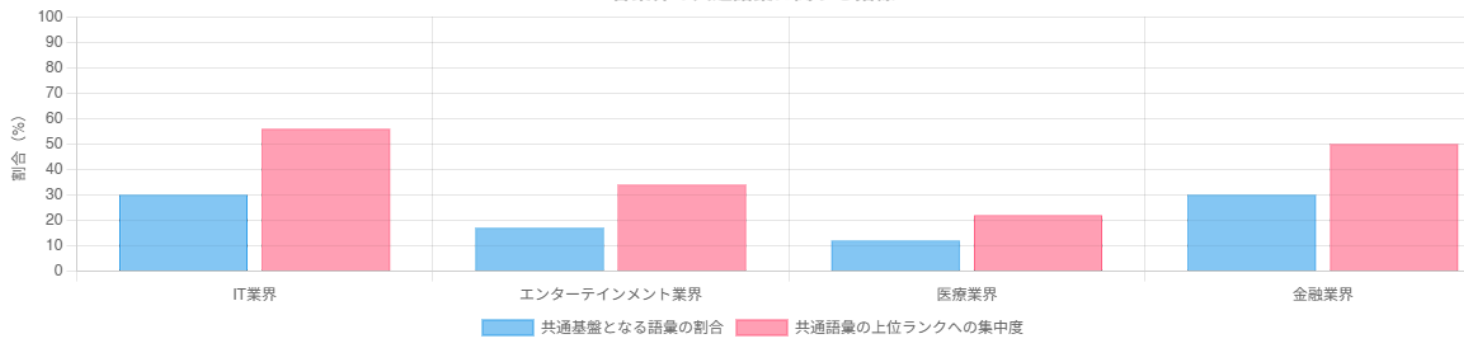
2. 分析結果

2.1 各業界の数値的特徴

分析対象とした4つの業界について、4つの指標の数値を以下の表にまとめる。

業界	共通基盤となる語彙の割合	共通語彙の上位ランクへの集中度	主要品詞（名詞）の構成比率	数値・記号を含む語彙の存在率
IT業界	約30%	約56%	85%以上	約4%
エンターテインメント業界	約17%	約34%	80%以上	約7%
医療業界	約12%	約22%	60%以上	約3%
金融業界	約30%	約50%	85%以上	約2%

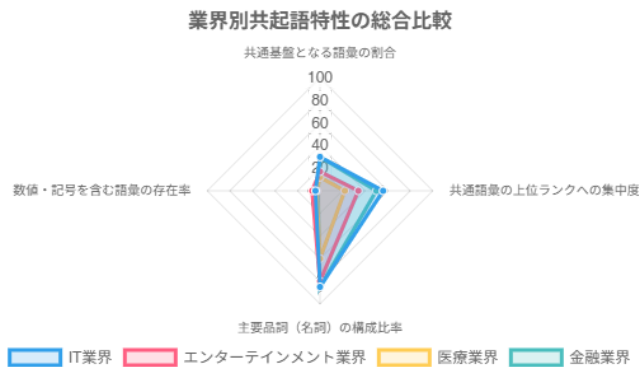
各業界の共通語彙に関する指標



各業界の特徴を簡潔に要約すると：

- **IT業界**：共通基盤となる語彙の割合が高く（約30%）、共通語彙の上位ランクへの集中度も高い（約56%）。名詞の構成比率が非常に高く（85%以上）、専門的な用語や概念が多く用いられていることが示唆される。
- **エンターテインメント業界**：共通基盤となる語彙の割合は中程度（約17%）で、共通語彙の上位ランクへの集中度も中程度（約34%）。数値・記号を含む語彙の存在率が最も高い（約7%）ことから、公開日や評価数値などの具体的情報が重視されている可能性がある。
- **医療業界**：共通基盤となる語彙の割合が最も低く（約12%）、共通語彙の上位ランクへの集中度も最も低い（約22%）。名詞の構成比率が他業界と比べて低い（60%以上）ことから、より多様な品詞が使用されている可能性がある。
- **金融業界**：IT業界と同様に共通基盤となる語彙の割合が高く（約30%）、共通語彙の上位ランクへの集中度も高い（約50%）。名詞の構成比率が非常に高い（85%以上）一方で、数値・記号を含む語彙の存在率は最も低い（約2%）特徴がある。

2.2 業界間の比較分析



業界間の比較から、以下のような顕著な差異や共通点が観察された：

共通基盤となる語彙の割合

IT業界と金融業界は共に約30%と高い数値を示し、共通基盤となる語彙が多いことがわかる。一方、医療業界は約12%と最も低く、サイト間での語彙の多様性が高いことが示唆される。エンターテインメント業界は約17%と中間的な値を示している。

共通語彙の上位ランクへの集中度

IT業界が約56%と最も高く、次いで金融業界が約50%と高い集中度を示している。医療業界は約22%と最も低く、共通語彙がランキング全体に分散している傾向がある。この指標も、IT・金融とエンターテインメント・医療で二極化の傾向が見られる。

主要品詞（名詞）の構成比率

IT業界と金融業界がともに85%以上と高い名詞比率を示し、エンターテインメント業界も80%以上と高い。一方、医療業界は60%以上と他業界と比較して明らかに低い値を示している。医療業界では、動詞や形容詞などの他の品詞が多く使用されている可能性がある。

数値・記号を含む語彙の存在率

エンターテインメント業界が約7%と最も高く、次いでIT業界が約4%、医療業界が約3%、金融業界が約2%という順になっている。エンターテインメント業界では、リリース日や評価点数などの数値情報が重視される傾向があると考えられる。意外にも金融業界で数値・記号を含む語彙の存在率が最も低いことは注目に値する。

業界グループの形成

分析結果から、IT業界と金融業界が多くの指標で類似した特性を示し、一つのグループを形成していることが観察される。一方、医療業界は多くの指標で他業界と異なる特徴を示している。エンターテインメント業界は、数値・記号を含む語彙の存在率が特徴的に高い点を除けば、中間的な特性を持つことがわかる。

3. 考察

3.1 数値データからの業界特性の推察

IT業界

IT業界では、共通基盤となる語彙の割合と共通語彙の上位ランクへの集中度が高く、またほとんどが名詞から構成されている。これは、IT業界のウェブコンテンツが特定の専門用語や技術概念を中心に構成されており、コンテンツの専門性と統一性が高いことを示している。検索ユーザーも特定

の専門知識や技術情報を求めている可能性が高く、上位サイトはそうした共通の情報ニーズに応える形でコンテンツを最適化していると考えられる。

エンターテインメント業界

エンターテインメント業界では、共通基盤となる語彙の割合と上位ランクへの集中度は中程度であり、業界内で一定の共通語彙が存在しつつも、サイトごとの特色や多様性も認められる。数値・記号を含む語彙の存在率が最も高いことは、公開日、評価点数、ランキング、シーズン番号など、具体的な数値情報がコンテンツの価値を高める要素となっていることを示唆している。エンターテインメントのユーザーは新しさや定量的評価を重視する傾向があると推察される。

医療業界

医療業界は4つの業界の中で最も特異な特徴を示している。共通基盤となる語彙の割合と上位ランクへの集中度が最も低く、名詞の構成比率も他業界と比較して明らかに低い。これは医療情報の多様性と複雑性を反映していると考えられる。医療関連の検索では、症状、原因、治療法、予防策など多角的な情報が求められ、それに応じてサイト間での語彙の多様性が高くなっていると推察される。また、名詞比率が低いことは、医療情報が状態や行動を説明する動詞や形容詞を多く含むこと、つまり「何か」だけでなく「どのように」「どのような状態か」といった説明的要素が重要であることを示唆している。

金融業界

金融業界はIT業界と多くの類似点を持ち、共通基盤となる語彙の割合が高く、名詞の構成比率も高い。これは金融商品や概念に関する専門用語が重要な役割を果たしていることを示している。一方で、数値・記号を含む語彙の存在率が最も低いことは、具体的な数値（金利や手数料など）よりも、商品の特徴や概念の説明に重点が置かれている可能性を示唆している。または、数値情報が表やグラフなど別の形式で提供されている可能性も考えられる。

3.2 品詞構成と情報の専門性・具体性

名詞比率の高さは、情報の専門性と概念的な密度を反映していると考えられる。IT業界と金融業界では名詞比率が85%以上と非常に高く、特定の製品名、技術用語、金融商品名などの専門用語を中心にコンテンツが構築されていることが示唆される。これらの業界では、「何について」という対象の明確化が重要であり、ユーザーも特定の概念や対象に関する情報を求めていると推察される。

一方、医療業界での名詞比率の相対的な低さ（60%以上）は、医療情報が名詞だけでなく、症状の説明（形容詞）や治療法・対処法（動詞）など、より多様な品詞を必要とすることを示している。医療情報の受け手は、単に疾患名などの名詞だけでなく、その状態や対処方法に関する詳細な説明を求めているのだろう。

数値・記号を含む語彙の存在率については、エンターテインメント業界が最も高く（約7%）、金融業界が最も低い（約2%）という意外な結果となった。エンターテインメント業界では、公開日、シーズン番号、評価点数など、コンテンツの新しさや評価を示す数値情報が重視されていると考えられる。金融業界で数値・記号を含む語彙の存在率が低いことは、可読性の高いコンテンツを重視する傾向があるか、または数値情報が表やグラフなど別の形式で提供されている可能性を示唆している。

3.3 業界横断的な共通傾向

4つの業界を比較した結果、いくつかの業界横断的な共通傾向が観察された：

- 名詞の優位性**：すべての業界で名詞が60%以上を占めており、ウェブコンテンツにおいては「何について」という対象の明確な提示が重視されている。
- 業界の専門性と語彙の共通性の相関**：IT業界と金融業界のように専門性の高い業界では、共通基盤となる語彙の割合が高い傾向がある。これは、専門分野において重要な概念や用語が存在し、それらが検索上位サイトに共通して含まれていることを示している。
- 情報の多様性と共通語彙の分散傾向**：医療業界のように情報の多角性が求められる分野では、共通語彙が検索結果全体に分散する傾向がある。これは、単一のトピックではなく、様々な側面からの情報提供が価値を持つことを示唆している。
- 数値情報の業界特性による重要度の差異**：数値・記号を含む語彙の存在率は業界によって大きく異なり、情報の具体性や信頼性の示し方が業界ごとに異なることを示している。

これらの共通傾向は、効果的なSEO戦略の構築において、業界特性を考慮することの重要性を示している。特に、共通基盤となる語彙の理解と、業界に適した情報提供の方法（名詞中心か多様な品詞を使うか、数値情報をどの程度含めるかなど）は、コンテンツ最適化の重要な要素となり得る。

4. 共起語分析から得られる学びと示唆

4.1 ウェブコンテンツの特性理解への貢献

共起語分析、特に本研究で用いた4つの指標は、各業界のウェブコンテンツの特性を理解する上で有効な手段であることが示された。具体的には：

- 共通基盤となる語彙の割合からは、業界内での情報の標準化度合いや、ユーザーが共通して求める情報の存在が理解できる。
- 共通語彙の上位ランクへの集中度からは、特定の重要概念が存在するか、あるいは多角的な情報提供が求められるかという業界特性が浮かび上がる。
- 主要品詞の構成比率からは、情報の提供方法（概念説明中心か、状態・行動の説明も重要か）について洞察が得られる。
- 数値・記号を含む語彙の存在率からは、具体的な数値情報の重要性や、情報の信頼性を示す方法に関する業界特性が理解できる。

これらの理解は、特定の業界向けのコンテンツを作成する際の指針となり、ユーザーの情報ニーズに合致したコンテンツ設計を可能にする。

4.2 SEO戦略への示唆

本研究の結果は、効果的なSEO戦略の構築に以下のような示唆を提供している：

- 業界特性に応じたキーワード戦略**：IT業界や金融業界のように共通基盤となる語彙の割合が高い業界では、その共通語彙を中心としたキーワード戦略が有効である可能性がある。一方、医療業界のように共通性が低い業界では、より多様なキーワードバリエーションを検討する必要がある。
- コンテンツの品詞構成の最適化**：業界に応じて、名詞中心の概念説明型コンテンツか、様々な品詞を用いた説明的コンテンツかを選択することが重要である。特に医療業界では、名詞だけでなく、症状や対処法を説明する多様な品詞を含めることが効果的と考えられる。
- 数値情報の戦略的活用**：エンターテインメント業界のように数値・記号を含む語彙が重要な業界では、公開日や評価点数などの数値情報を積極的に活用することがSEOに寄与する可能性がある。

- ・**上位ランクへの重要語彙の集中**：IT業界や金融業界のように上位ランクへの集中度が高い業界では、特に重要な共通語彙をコンテンツの冒頭部分や見出しに配置することが効果的かもしれない。

これらの示唆は、単に検索エンジンのアルゴリズムに対応するだけでなく、各業界のユーザーが実際に求める情報の提供方法を最適化することにつながる。

4.3 業界ごとの情報流通の特徴把握

共起語分析の結果からは、各業界における情報流通の特徴も読み取ることができる：

- ・**IT業界**：専門性の高い技術情報が中心であり、特定の概念や技術用語を軸とした情報提供が行われている。情報の標準化度合いが高く、業界内での共通言語が形成されている。
- ・**エンターテインメント業界**：新しさや評価を数値化した情報が重要視され、コンテンツの多様性と具体性のバランスが特徴的である。トレンドやランキングなどの変動する情報が重要な役割を果たしている可能性がある。
- ・**医療業界**：情報の多角性と説明の詳細さが特徴的であり、単一の概念だけでなく、その状態や対処方法など多面的な情報提供が求められている。専門用語だけでなく、一般の人にもわかりやすい説明が重視されている可能性がある。
- ・**金融業界**：専門性の高い概念説明が中心であり、用語の標準化が進んでいる。具体的な数値よりも商品特性や概念の説明に重点が置かれている可能性がある。

これらの特徴把握は、各業界における効果的な情報発信の戦略立案に役立つとともに、業界の情報流通の特性に関する学術的理解を深めることに貢献する。

5. 結論

5.1 研究のまとめ

本研究では、IT、エンターテインメント、医療、金融という4つの主要業界における検索上位サイトの共起語分析を通じて、各業界のウェブコンテンツ特性を明らかにすることを目的とした。共通基盤となる語彙の割合、共通語彙の上位ランクへの集中度、主要品詞（名詞）の構成比率、数値・記号を含む語彙の存在率という4つの指標を用いた分析の結果、各業界には明確な特徴があることが示された。

IT業界と金融業界は、共通基盤となる語彙の割合と名詞の構成比率が高く、専門的な用語や概念を中心としたコンテンツ構成が特徴的である。エンターテインメント業界は、数値・記号を含む語彙の存在率が高く、具体的な数値情報が重要な役割を果たしている。医療業界は、共通基盤となる語彙の割合が低く、名詞の構成比率も他業界と比較して低いことから、多様な情報と説明的要素の重要性が示唆された。

これらの発見は、効果的なウェブコンテンツ戦略の構築やSEO対策において、業界特性を考慮することの重要性を示している。また、共起語分析が業界ごとの情報流通の特徴を理解する上で有効な手法であることも確認された。

5.2 研究の限界と今後の課題

本研究にはいくつかの限界があり、それに伴う今後の課題が存在する：

- ・**業界の代表性**：本研究では4つの業界に焦点を当てたが、より多様な業界を分析することで、より包括的な理解が得られる可能性がある。
- ・**時間的変動の考慮**：検索上位サイトは時間とともに変動する可能性があり、長期的な観察を通じて、共起語特性の時間的変化を分析することも重要である。
- ・**地域差の検討**：言語や地域によって共起語の特性が異なる可能性があり、異なる言語や地域での分析も今後の課題である。
- ・**因果関係の解明**：共起語の特性がSEOランキングにどの程度影響を与えているのか、その因果関係を明らかにすることも重要な課題である。
- ・**ユーザー行動との関連性**：共起語の特性とユーザーの行動（クリック率、滞在時間など）との関連性を分析することで、より実用的な知見が得られる可能性がある。

これらの課題に取り組むことで、共起語分析に基づくウェブコンテンツ最適化の理解をさらに深めることができるだろう。

5.3 実務への応用

本研究の成果は、以下のような実務への応用が期待される：

- ・**業界特性を考慮したコンテンツ戦略の立案**：各業界の特性に合わせたコンテンツ構成（専門用語の使用度合い、説明の詳細さ、数値情報の活用など）を検討することが可能になる。
- ・**効果的なキーワード戦略の構築**：共通基盤となる語彙の理解を通じて、各業界で重要視されるキーワードを特定し、効果的なSEO戦略を構築することができる。
- ・**競合分析の精緻化**：共起語分析の手法を用いて、競合サイトとの差別化ポイントや、業界内での位置づけを明確にすることが可能になる。
- ・**ユーザーニーズの理解**：共起語の特性から、各業界のユーザーが求める情報の性質（専門性、多角性、具体性など）を理解し、それに応じたコンテンツ提供を行うことができる。

これらの応用を通じて、より効果的なウェブコンテンツ戦略の構築とSEO対策の実現に貢献することが期待される。

論文構成案（骨子）

本研究の成果を論文としてまとめる場合、以下のような章立てを提案する：

1. 序論
 - ・研究背景：SEOにおける共起語分析の重要性と業界特性の関連
 - ・研究目的：異なる業界における共起語特性の比較分析と示唆の導出
 - ・分析対象と手法：4業界、4指標による分析アプローチの説明
2. 先行研究
 - ・共起語分析に関する既存研究
 - ・SEOと業界特性に関する研究
 - ・ウェブコンテンツの品詞構成に関する研究
3. 研究方法
 - ・分析対象となる業界とキーワードの選定

- データ収集方法：検索上位サイトの特定と共起語リストの作成
- 分析指標の定義と算出方法

4. 分析結果

- 各業界の共起語特性：4指標による数値的特徴
- 業界間の比較分析：類似点と相違点
- 業界グループの形成：特性に基づく業界の分類

5. 考察

- 数値データからの業界特性の推察
- 品詞構成と情報の専門性・具体性の関連
- 業界横断的な共通傾向の意味するもの

6. 共起語分析の意義と示唆

- ウェブコンテンツの特性理解への貢献
- SEO戦略への示唆
- 業界ごとの情報流通の特徴把握

7. 結論

- 研究のまとめ
- 研究の限界と今後の課題
- 実務への応用可能性

このような構成により、研究の背景から結論までを論理的に展開し、共起語分析の業界特性に関する理解を深めるとともに、実務への応用可能性を示すことができる。

参考文献

1. Loet Leydesdorff & Liwen Vaughan (2006). *Co-Occurrence Matrices and Their Applications in Information Science: Extending ACA to the Web Environment*. *Journal of the American Society for Information Science and Technology*, 57(12), 1616–1628. [DOI](#)
2. Ying Ding, Gobinda Chowdhury & Schubert Foo (2001). *Organising Keywords in a Web Search Environment: A Methodology Based on Co-Word Analysis*. In *Proceedings of the 8th International ISKO Conference*. [PDF](#)
3. 玉岡 賀津雄 (2016). 「共起表現研究のためのコーパス検索入門」 『レキシコンフォーラム』 7, 239–264. [PDF](#)