検索から生成へ：生成AI時代における検索技術の再定義

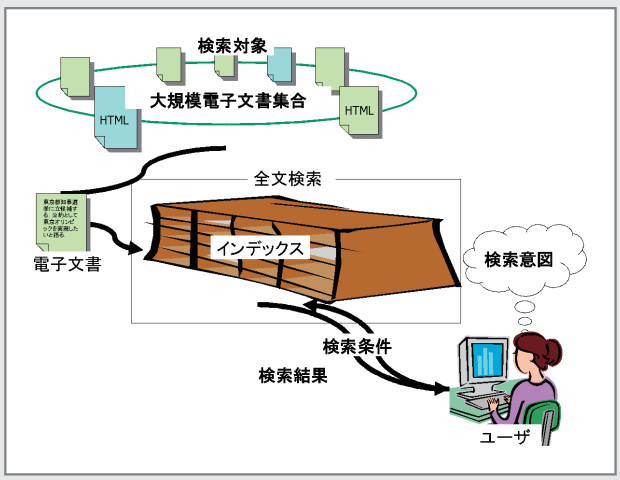
東京理科大学 創域理工学部 情報計算科学科 准教授 植松（うえまつ） 幸生（ゆきお）

■ はじめに

本稿では，筆者がこれまでに取り組んできた検索システムの研究開発の経験を踏まえ，生成AIの登場によって情報検索の構造や意味がどのように変化したのかを分析する．生成AIの代表例として大規模言語モデル（LLM）が挙げられるが，従来の情報検索（Information Retrieval: IR）とは異なるアプローチを取りながらも，ユーザの情報ニーズに応えるという本質的な目的は共通である．本稿では，IRと生成AIの類似点と相違点を明らかにし，両者の融合によって今後の検索技術がどのように再定義されるべきかについて論じる．

■ 第1章：検索システムの本質とは何か？

検索システムは，ユーザが明示的に入力したクエリに対して，あらかじめ構築されたインデックスから関連性の高い文書を順序付け（ランキング）し提示するシステムである．その中核は「ユーザ意図の理解」と「情報の構造化」にある．ユーザ意図の理解とは，入力されたクエリがユーザのどのような情報ニーズに対応しているのかを正確に解釈することである．一方，情報の構造化とは，大量に存在する文書データを効率的に検索可能な形式に整理・保存することを指し，インデックスと呼ばれるデータ構造で実現されている．



【図1】従来の検索システム（キーワード検索）

【図1】に従来の検索システム（キーワード検索）のおおまかな構成を示す．従来の検索システムは，ユーザがコンピュータに検索条件（クエリ）を入力すると，その条件に基づき大規模な電子文書集合の中から関連性が高いと推測される文書を，全文検索のインデックスを用いて素早く探し出す．電子文書は事前に処理され，検索対象としてHTML文書やその他形式の文書がインデックス化されている．このインデックスは巨大な書棚のように情報が整理されており，検索効率を高める役割を担う．

従来の検索システムでは，TF-IDF，BM25，Learning to Rankなどのランキング方法やランキングのための重みづけ手法が主に用いられてきた．例えば最も代表的なTF-IDFは，文書中のキーワードの重要度を，文書内でのキーワードの頻度（TF: term frequency）と，全体の文書集合におけるそのキーワードの希少性（IDF: inverse document frequency）に基づいて数値化する方法である．BM25はTF-IDFをさらに発展させた手法で，文書長やキーワードの出現頻度の飽和効果を考慮し，より実用的なランキングを提供する．Learning to Rankは，機械学習を活用して，ユーザが評価した過去のデータからランキングモデルを学習することで，検索精度を動的に向上させる技術である．検索ランキングは，入力されたクエリqと検索対象となる文書dとのマッチングによって決定されるが，前述したTF-IDFを利用した類似度のような単なるキーワードマッチングにとどまらず，文書の信頼性やユーザ行動履歴などのクエリ非依存情報も活用される．例えば，「東京都 墨田区役所」と検索した場合，ユーザが求める情報が墨田区役所の公式ページである可能性が高いため，単に「東京都 墨田区役所」のような文字列が出現するだけのページではなく，公式ページが最上位に提示される．この際，公式ページであることを示すページランクやドメインの信頼性といった情報がランキングに重要な役割を果たす．ページランクは，多くのページからリンクされたページは信頼されるページであることを重みづけしたもので，ドメイン

の信頼性とは，例えば墨田区のページ（https://www.city.sumida.lg.jp/）のように lg.jp のような信頼できる top level domain から提供されているかが重みづけの対象になる．

検索システムにおけるもう一つ重要な要素は，評価指標による性能管理である．特に重要視されるのが適合率（Precision）と再現率（Recall）であり，この二つはトレードオフの関係にある．適合率は，検索結果として提示された文書のうち，実際にユーザが求めている情報に適合している文書の割合を示す．一方，再現率はユーザが本来求めている文書全体の中で，検索結果として提示された適合文書の割合を示す．大量の適合文書を返せば Recall が上がる傾向になり，逆に Precision は適合している文書数が限られていると下がることになる．例えば，検索結果の上位 10 件について評価を行う場合は，Precision@10 という上位 10 件のみを評価対象とした指標を使用し，システムの精度を比較評価する．

検索システムの設計者は，こうした評価指標を参考にしながら，検索意図の理解を深め，ユーザが最も必要としている情報を迅速かつ的確に提示できるような検索アルゴリズムやランキングモデルを構築することが求められている．

従来の検索システム（キーワード検索）の構造をさらに具体的に説明すると，ユーザは主にキーワード形式で検索条件を入力し，検索システムはその条件に基づいて，インデックス化された文書集合から最も適合度が高いと推定される文書をランキング形式で順序付けして提示する．検索結果ページでは，文書のリンクとともに，ユーザが入力したキーワードが文書内のどの部分でどのように使用されているかを示す「KWIC: Keyword in context」表示も提供される．

しかしながら，こうした従来の検索システムはあくまでインデックス化された文書集合内のマッチングに基づくリンク提示と，キーワードの出現位置の強調表示にとどまっており，検索結果の内容を深く解釈したり，ユーザが本質的に求めている具体的な回答や情報をインデックスに保存されている情報から加工・再構成して提示する機能はなかった．この制限が，後の生成 AI を用いた検索システムの登場と進化を促す一つの要因となったのである．

■ 第 2 章：生成 AI の登場と情報検索に与えた衝撃

大規模言語モデル（LLM）による生成 AI の登場は，従来の検索システムの枠組みに大きな変革をもたらした．ChatGPT（OpenAI 社），Claude（Anthropic 社），Gemini（Google 社）などに代表される生成 AI の基盤モデル（Foundation model）は，従来の検索システムのような単なる検索結果のリスト提示から脱却し，自然言語による応答と生成という新たな情報提供手法を導入した．この劇的な変化は，ユーザ体験（UX: User Experience）に直接的かつ大きな影響を与えている．

従来の情報検索では，ユーザが特定の情報を「調べる」ために適切なキーワードを用い，検索結果の中から関連する情報が記載された文書を自ら選択する必要があった．一方で，生成 AI の登場以降，ユーザが情報を「質問する」形式へと変化したことで，具体的な文書の選択なしに直接的に情報を得られるようになった．これは，ユーザが求めている情報そのものを AI が理解し，意味単位で情報を生成・提供するという新しい検索結果の提示パラダイムへの移行を意味していると考える．

実際の比較例として，**【図 2】**に「小学生でも分かる民主主義の説明」というクエリに対する従来の検索システムから出力される SERP（Search Engine Result Page）を示す．従来の検索システム（例えば Google など）は，このクエリに対して，小学生でも理解できるような民主主義の説明を掲載したホームページをリスト形式で提示する．ユーザはその中から適切な情報を探し，選択する必要がある．

これに対して**【図 3】**に生成 AI の一つである ChatGPT（GPT4o）の出力を示す．ChatGPT は，クエリに対して簡潔かつ分かりやすい説明を生成するだけでなく，具体的で身近な例え話を提供することにより，より直感的に理解を促すことを目指した出力になっている．

また，この事例とは異なるが質問が曖昧であったり，ユーザがうまく質問を表現できてない場合でも，チャット形式でユーザがやり取りを重ねることにより，本来の検索意図を掘り下げ，最終的にはユーザが求めている具体的かつ明確な回答へと誘導することが可能である．

このように生成 AI の導入によって，ユーザが得られる情報の質や速度，そして利便性が飛躍的に向上した．これまでユーザが行ってきた情報の取捨選択や判

【図 2】従来の検索システムの検索結果例 (小学生でも分かる民主主義の説明)

【図 3】生成 AI の出力例（小学生でも分かる民主主義の説明）

【図 4】Felo Search の検索結果例（小学生でも分かる民主主義の説明）

断プロセスを AI が支援，あるいは一部代替することで，検索システムの価値や役割そのものが再定義されつつある．さらに，生成 AI の応答が精度や正確性を高めるにつれて，ユーザの信頼感や満足度も大幅に向上し，情報検索の新たな標準として広く受け入れられ始めている．余談ではあるが，この情報収集や検索に特化した生成 AI を利用したサービスを多くのベンチャー企業が提供している．(ex. Felo Search や Perplexity)

例えば，【図 4】に Felo Search で小学生でも分かる民主主義の説明を示す．

Felo Search は入力されたクエリに対して，Web 検索に適したクエリを内部的に作成し，検索した結果から答えを生成する．ChatGPT との違いは，結果を出力する際に利用したプロンプトの過程が表示されたり，情報ソースが画面の右側に提供される．これにより信頼できる情報ソースから得られた結果なのかどうかが判断することができる．

今後は，生成 AI と従来型の検索システムが相互補完的に運用されることで，それぞれの利点を活かしたハイブリッド型の検索環境が形成される可能性も高いだろう．AI 技術の進化とともに，情報検索はさらなる進化を遂げていくことが予想される．

■ 第 3 章 : 何が“変わらない”のか？

生成 AI が登場したとはいえ，情報検索における根本的な要請は変わっていない．第一に，ユーザは依然として「自らの問いに対する信頼できる答え」を求めている．これは情報検索（IR）におけるユーザ意図の理解と，情報の適切なマッチング（クエリ依存の重要度），さらに情報提供元の信頼性や重要性（クエリ非依存の重要度）が現在でも継続していることを示している．つまり，ユーザが得たいのは単なる情報ではなく，正確で信頼性が高く，明確に理解可能な情報であることに変化はない．

第二に，生成 AI においても情報検索技術はバックエンドで重要な役割を果たしていることが挙げられる．その代表的な例が RAG（Retrieval-Augmented Generation）と呼ばれる技術である．この技術では，ユーザが入力したクエリを基にして，まずバックエンドで従来型の検索システムを用いて関連文書を検索し，それらの文書をプロンプトの入力として利用することで，生成される回答の新鮮度，あるドメインに特化した情報の取得，正確性や信頼性を高めることが出来る．

【図 5】に生成 AI を用いた検索システムの構成を示す．RAG では，ユーザが生成 AI（LLM）に対してクエリを投げかけると，LLM は tools（function calling 機能）を利用し，どのデータベースにアクセスすべきか，あるいは自身が持つ既存知識だけで回答可能かを自動で判断する．このプロセスにおいて，例えば特定企業に関する質問が入力された場合，LLM はその企業に関する情報を特化して保持している企業データベースや，WebAPI を通じて企業のホームページに対して問い合わせを行い，その情報を取得し，その情報を入力として回答を生成する．このプロセスにおいては，実質的に従来型の検索システムと同様の検索処理がバックエンドで機能している．例を挙げると，「〇

〇株式会社の決算情報から費用として大きい費目の上位５件を教えて」といった具体的なクエリの場合，LLM は企業 DB に対して適切なクエリを作成し，実際に検索システムを用いて該当データを取得し，それを基にユーザへの回答を生成する．このようなクエリは単なる情報生成ではなく，正確なデータ取得が不可欠であり，まさに従来の IR システムが果たしてきた機能が求められる．

また，従来の情報検索で培われてきたユーザ行動ログの解析技術やランキング評価技術は，生成 AI が提供する回答の評価や精度改善にも積極的に利用されている．例えば，生成された回答に対するユーザのクリック率や満足度，さらにフォローアップの質問や修正要求といったログ情報を活用することで，生成 AI の性能を継続的に評価し改善することが可能となる．

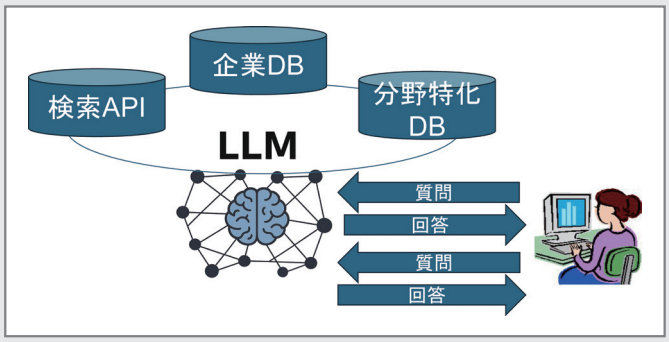
最後に速度である．生成 AI に質問してから回答までの時間は従来の検索システムの時間よりも比較的長い．また，function calling 機能から Web 検索の API を呼び出したりする場合は，結局検索システムの速度が最終的な検索結果の出力時間と直結するため，依然として重要な要素と言える．

これらの技術的な基盤としての情報検索の価値は，生成 AI の登場以降も全く変わらず，むしろその重要性は増していると言える．

■ 第４章：何が“変わった”のか？

一方で，生成 AI は検索技術に対していくつかの本質的な変化をもたらした．その中でも特に顕著なのが「インターフェースの変化」である．従来の検索システムでは，ユーザがキーワードを入力すると，関連する多数の文書リンクがリスト形式で提示される．ユーザはそれらの文書を一つずつ確認し，自分の検索意図に最も合致するものを選択する必要があった．このプロセスは検索に関する負担や時間を伴い，特にユーザが明確なクエリを設定できない場合や複雑な情報を求めている場合には，そのコストが大きかった．

これに対して生成 AI は，単一の自然言語応答をユーザに提供することで，情報を精査するプロセスを劇的に簡素化している．例えば，「日本の民主主義の歴史について簡単に教えて」という質問に対しても，従来はユーザ自身が複数の文書を比較して情報を取捨選択する必要があったが，生成 AI は一つのまとまった説明を即座に提示できる．これにより，情報の精査にかかる認知的負担や時間的コストが大幅に軽減された．



【図 5】Retrieval Augmented Generation（RAG）を利用した検索の模式図

第二の大きな変化として挙げられるのが，ユーザが求める情報の質の変化である．従来は，検索結果の信頼性や正確性が高い情報が求められていた．しかしながら，生成 AI の普及に伴いユーザが求めるものは「納得できる説明」に変化しつつある．これは，生成 AI の応答が厳密な事実に基づくものであることはあくまで前提として，意味的整合性や言語的な自然さを保つことで，ユーザが受け入れやすい情報を生成することに起因している．たとえば，政治や経済の複雑な概念を質問された際に，生成 AI は精緻な事実の情報よりも理解しやすく親しみやすい説明を重視して応答を生成する傾向にある．

具体的な例を挙げると，ユーザが「インフレがなぜ起きるのか？」という質問をすると，従来の検索システムは複数の経済学者の記事や政府統計のリンクを提示する．一方で，生成 AI は具体的な事例や比喩を用いてわかりやすい説明を作り出し，経済学の知識が浅いユーザにも直感的に理解可能な応答を提供できる．ここで重要なのは，理解のしやすさや説明の説得力がより評価されるようになっているという点である．

第三に，情報源の透明性や出典表示の問題も変化している．従来の検索システムでは，文書が明確に表示されるため，ユーザは情報の出典を容易に確認できたが，生成 AI では回答が新たに生成されるため，元となる情報源が曖昧になることもある．例えば，医学的アドバイスや法律相談といった専門性の高い領域では，この出典の曖昧性がユーザの意思決定に影響を与える可能性がある．

このような変化に伴い，IR（情報検索）で伝統的に重視されてきた「信頼性」や「再現性」の概念が再定義される必要が出てきている．具体的には，生成 AI の応答に対して，その情報の出典を明確に表示する技術的な仕組みや，生成された情報の信頼性を評価・保証する新たな基準や評価手法の導入が求められるよう

になっている．

以上のように，生成 AI は検索システムのインターフェースや求められる情報の性質，そして信頼性評価の方法に至るまで，情報検索技術の多くの側面で重要かつ大きな変化を引き起こしている．

■ 第 5 章：検索と生成の融合

現在，検索と生成を融合させたハイブリッド型の情報検索が注目を集めている．このような融合型の検索技術の最も顕著な例は RAG（Retrieval-Augmented Generation）である．RAG は，前章までで述べた通り検索技術と生成 AI の利点を組み合わせた手法であり，ユーザが入力したクエリに基づき関連性の高い文書を検索システムを用いて抽出する．その後，生成 AI がこれらの抽出された文書を入力として自然言語で回答を生成する仕組みである．

従来から検索システムの評価や評価指標を活発に議論している国際会議として TREC（Text REtrieval Conference）がある．TREC の RAG トラックでは，従来の情報検索と生成技術の接点と違いを明確に検証しようという動きが活発化している．

TREC の RAG トラックは，設計当初から 2 つの段階的なタスクに分かれて構成されている．第 1 タスクは，ユーザクエリに対して関連する文書を検索するというもので，これは従来の TREC の ad-hoc トラックとほぼ同様の構成である．評価指標としても nDCG や Recall といった従来の検索精度を測定する評価指標が用いられており，ここでは検索システム単体の文書取得能力が評価される．

第 2 タスクでは，取得した文書を元に LLM が自然言語で応答を生成するタスクである．この生成タスクは，単に文書を並べるのではなく，そこから必要な情報を抽出・要約し，文脈に沿った回答を構築する能力が問われる．生成された応答は，人間による評価（Human Evaluation）により「情報の正確さ」「質問との関連性」「自然さ」などの観点で評価される．従来の ad-hoc タスクが文書単位の関連性にフォーカスしていたのに対し，RAG タスクは意味単位での情報の再構成と提示を目的としている点で大きく異なる．

たとえば，「アメリカ独立戦争の要因は何か？」という質問に対し，ad-hoc タスクでは Wikipedia や正式な歴史書等の該当文書が提示されることが求められる．一方，RAG タスクでは，複数の文書から情報を取り出し，「税制への反発と政治的自由への要求が主因である」といった直接的かつ要約された応答を生成することが求められる．TREC の 2025 年では，これを踏襲しつつ，R（Retrieval task），AG（Augmented Generation task）それに RAG task という 3 種類のタスクを実施することがアナウンスされている．

このように TREC RAG トラックでは，従来の情報検索の文脈を継承しつつも，生成という新しいパラダイムとの融合を評価するための二層構造が採用されている．このアプローチは，検索と生成が別々の技術としてではなく，補完的に機能する設計が求められているという現代の情報検索における潮流を反映している．

今後は，TREC RAG トラックのような国際的なベンチマークを通じて，検索と生成がいかに融合し，互いを高め合うのかという観点から技術評価が進められていくだろう．これにより，検索の信頼性と生成の柔軟性を両立する高度な情報検索システムの実現が期待されている．

さらに，この融合型検索の普及と発展には，新たな課題への取り組みが不可欠である．特に重要な課題として，クエリの意味理解（Semantic Understanding），情報の文脈的接続（Contextual Integration），生成結果の検証可能性（Verifiability）の確保が挙げられる．例えば，ユーザが「2023 年の経済予測」といった曖昧で広範なクエリを入力した場合でも，検索システムはその意図を正確に解釈し，適切な情報源を識別して文脈的に関連する情報を集約・生成する能力が求められる．また，生成された回答が具体的にどの文書やデータに基づいているのかをユーザに明示することが，情報の信頼性や透明性を確保するために極めて重要となる．

今後，生成 AI と IR 技術の融合が進展するにつれ，よりユーザの情報ニーズに対して高精度かつ柔軟に対応できる新たな検索システムの構築が期待される．このような新たなシステムは，情報の正確性，信頼性，そして理解の容易さという三つの要素を高度に統合し，ユーザ体験を一層向上させる可能性を秘めている．

■ 第 6 章：これからの検索技術者に求められる視点

生成 AI 時代における検索技術者には，従来型の検索アルゴリズムやインデクシングの最適化といった技術的スキルに加えて，モデル構築の知識，ユーザ行動やニーズへの深い理解，そして UX 設計に関する知見が新たに求められている．これまで検索技術者の役割

は主に検索精度の向上や処理速度の最適化といった技術的課題の解決に集中していたが，生成 AI が普及するにつれ，プロンプト設計やユーザの文脈把握，生成結果の信頼性評価など，より広範な課題への対応能力が重要視されている．

具体的には，プロンプト設計では，どのようなクエリが適切な回答を引き出すかを理解し，ユーザが求めている情報を最適な形で引き出す工夫が必要となる．例えば，医療情報に関する問い合わせを考えると，「頭痛の治し方」ではなく，「慢性的な片頭痛の効果的な治療方法」など具体的で適切なプロンプト設計が求められる．また，ユーザ文脈の抽出に関しては，ユーザが直前に行った検索行動やその時間帯，使用デバイス，地理的位置情報といったコンテキストを活用して，個々のニーズに適合した情報提供を実現する必要がある．

さらに，生成内容の信頼性評価も重要な役割を担う．従来の評価指標である BLEU や ROUGE といった表層的なテキスト一致指標だけではなく，「納得度」や「役立ち度」といった主観的評価指標の導入も重要である．また，これを自動化するアプローチとして llm as a judge 等が評価方法の一つとして注目されている．

例えば，法律や医療，経済などの専門的な情報に関しては，単に生成された文章の流暢さや情報量だけでなく，専門家やユーザ自身による評価を通じて，本当に信頼に値する情報かどうかを慎重に判断する必要がある．このためには，生成された応答がどの文書やデータを参照しているかを透明化する仕組みや，ユーザが情報の真偽や出典を容易に確認できるような設計が重要となる．

また，検索技術と生成 AI を対立的に捉えるのではなく，互いを補完する存在として捉え，それぞれの長所を活かしたシステム設計や実装を行う能力も求められている．検索の正確性や迅速性を生成 AI の柔軟性やユーザフレンドリーなインターフェースと融合させることで，UX の劇的な向上が期待できる．こうした融合的な視点を持った人材の育成が，情報社会のさらなる発展を促進する鍵となる．

結論として，生成 AI 時代の検索技術者には，従来の検索アルゴリズム開発に加え，ユーザ行動理解，プロンプト設計，生成内容の評価手法の開発など，多面的かつ総合的な能力が求められる．これらの能力を兼ね備えた人材こそが，今後の情報検索技術の進化と発展を牽引していくことになる．

■ おわりに

検索から生成への進化は，単なる技術の置き換えにとどまらず，私たちが情報とどう向き合い，どのように理解・活用するかという根本的な姿勢そのものを変革しつつある．事実，私自身も Google 検索で情報を自ら探す回数が，近年格段に減ったと感じている．

この変化は，単に「検索システム」から「対話型 AI」へのインターフェースの転換にとどまらない．ユーザの役割は，単に情報を「選ぶ」ことから，AI と「対話しながら構築する」プロセスへと移行している．

筆者は，長年にわたり情報検索（IR）のシステム開発に関わってきた立場から，この変化を大きな行動変容と捉えると同時に，従来の情報検索の基盤技術がこれまで以上に重要になったと考えている．たとえば，生成 AI と検索を連携する場合には，生成 AI に正確な情報を入力しなければ，正しい出力は得られない．また，生成 AI が回答を生成するまでの速度も，検索の速度が影響を及ぼす．したがって，情報検索の「精度」と「速度」の向上は，今後さらに重要になるだろう．

本稿では，第 1 章から第 6 章を通じて，従来の検索技術の原理，生成 AI のインパクト，情報検索と生成 AI の接点，そして技術者に求められる視点の変化について俯瞰してきた．その中で明らかになったのは，今後の検索システムは「情報源の信頼性／透明性」「理解しやすさ」「文脈適合性」「生成までの速度」など，多様な価値基準のバランスの上に成り立つということである．検索と生成は，互いを補完し合うものであり，それらを効果的に融合させるための技術的な設計が，我々に問われている．

これからの検索技術者には，検索技術と生成技術の両面を理解したうえで，ユーザと情報提供者それぞれの意図を読み解き，対話を設計し，意味の構築を支援できる新たなスキルセットが求められる．そして，情報の信頼性を維持しながら，AI が生成するコンテンツの価値を最大化するための評価方法やガイドラインの整備も，今後の大きな課題となる．本稿が，検索技術の過去・現在・未来を俯瞰する一助となり，生成 AI 時代における情報検索の可能性と責任について考える契機となれば幸いである．技術が進化する今こそ，私たち一人ひとりが情報とどう向き合うかが，これまで以上に重要になる．