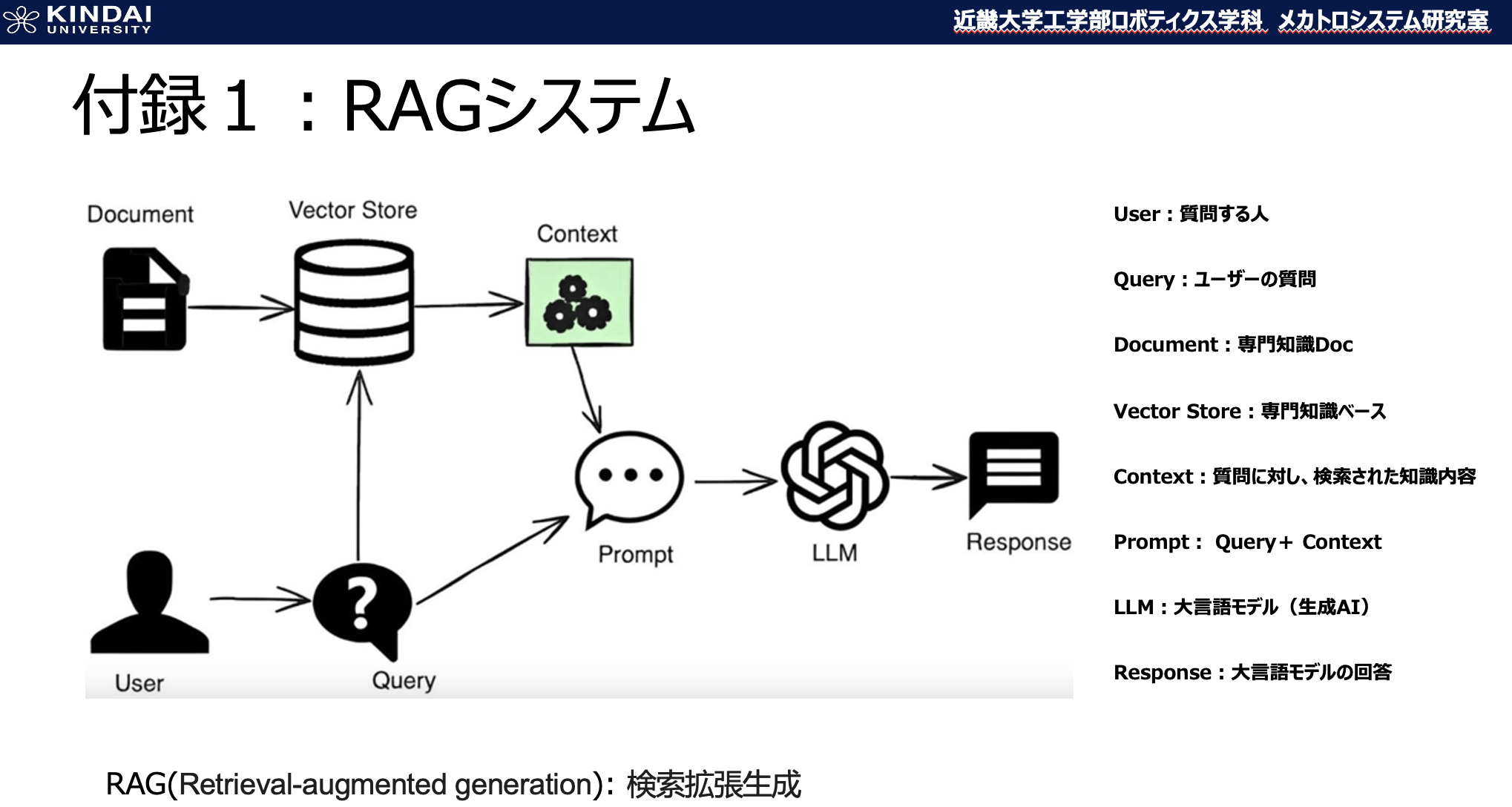
# 研究説明

## 1. RAGとは何か

RAG（Retrieval Augmented Generation）は、大型言語モデル（LLM）が持つ事前知識の限界を補完するために、外部知識源から関連情報を検索し、それを基に応答を生成する手法である。具体的なプロセスは以下の通り：  
  
- ユーザーの質問を入力  
- 検索モジュールが外部知識源から関連文書を取得  
- 取得文書と質問をLLMへ入力  
- より正確かつ最新の回答を生成  
  
このアプローチにより、専門的な内容や最新情報に関する質問にも高精度に対応可能となる。ただし、従来の研究は主に英語環境で行われてきた。

## 2. 多言語RAGの課題

本研究が対象とするのは、「非英語の質問（例：日本語）に対し、主に英語の知識源から検索を行い、日本語で正確に回答する」クロスリンガルRAGである。現状では以下のような課題がある：  
  
① 検索戦略に関する課題  
  
- 翻訳式検索（tRAG）：質問を英語に翻訳してから検索する方式だが、翻訳誤りにより重要な情報を見逃すリスクがある。  
- 多言語検索（MultiRAG）：複数言語から直接検索可能だが、取得文書に言語の混在があると生成精度に悪影響を与える。  
- CrossRAG：検索結果を一つの言語に統一してから生成する方式が提案されているが、設計には細心の注意が必要。  
  
② 回答言語の整合性  
  
正しい回答が生成されたとしても、「日本語の質問に対して英語で回答される」「一部に英語が混じる」といったコードスイッチが頻発する。これはユーザーの混乱を招くため、明確な言語制御が不可欠である。  
  
③ クロスリンガル推論の困難性  
  
XRAGの研究では、複数言語から文書を取得しても、モデルがそれらを正確に統合・推論できない例が多く報告されている。生成そのものよりも、「検索→整合→推論」プロセスの方が本質的に難しいことが明らかになっている。

## 3. 提案手法： LangBridge

本研究では以下の改良を行うシステム「UniTransRAG（またはLangBridge）」を提案する：  
  
1. 二重検索チャンネル：日本語と英語の両方で検索を実行し、情報の網羅性と精度を確保する。  
2. 取得文書の翻訳統一：検索で得られた多言語文書を一旦すべて英語または日本語に統一してから生成フェーズに送る。  
3. プロンプトによる言語制御：「日本語で回答せよ」と明示的に指示し、回答例を与えることで言語の混用を抑制する。  
4. 生成後の自己検証：生成された日本語回答を英語に再翻訳し、元の文書との整合性を確認するプロセスを導入予定。  
5. 言語一致性の評価：回答文中に日本語以外の単語が含まれていないかを自動的に判定し、言語混用の頻度を評価指標とする。

## 4. 実験および評価方法

- 使用データセット：MKQA（26言語対応）、XOR TyDi QA（非英語質問＋英語知識ベース）、必要に応じてXRAGベンチマークも併用。  
- 知識源：多言語Wikipediaおよび英語Wikipediaを組み合わせ、方式ごとに使い分ける。  
- 比較対象モデル：  
 - 翻訳式RAG（tRAG）  
 - 多言語RAG（MultiRAG）  
 - 提案手法（統合型CrossRAG）  
  
- 評価指標：  
 - 検索精度（Recall@K）  
 - 回答正確度（EM、F1、BLEU）  
 - 言語一致性：回答中に日本語以外が混入しない比率  
 - コードスイッチの発生頻度などの定性的評価  
  
- XRAGベンチマーク：単言語検索／多言語検索の両方に対応し、生成言語の一貫性とクロスリンガル推論能力を検証する。

## 5. 期待される成果

- 言語一致性と正答率の向上により、日本語話者にとって実用的なクロスリンガルRAGを構築  
- 高リソースおよび低リソース言語の双方での有効性を確認し、多言語対応の妥当性を証明  
-再現可能なベースラインモデルとしてコード付きで公開する