

OpenAlex 再構築のロールアウトタイムライン発表

概要

- OpenAlex は今後 5 か月間で新しいコードベースへ移行されます。
- スキーマ(データ構造)は変更されませんが、約 5%のデータが変化し、5,000 万件以上の新しい論文・研究成果が追加されます。

変更の背景

- OpenAlex は Microsoft Academic Graph の終了を受けて急いで作られたため、コードが複雑で保守が困難でした。
- Unpaywall との統合が不十分で、バグ修正や新機能追加が非常に困難。
- 新しいデータソース(例: DataCite)やエンティティ(例: 助成金)の追加がほぼ不可能。

解決策

- Unpaywall と OpenAlex のコードベースを統合し、Apache Spark (Databricks 上)で再構築。
- より現代的で保守性・速度に優れた基盤へ刷新。

主な変更点

- **5,000 万件以上の新規追加**
DataCite や機関リポジトリなどから大量の新しい研究成果を追加。
- **Unpaywall と OpenAlex の整合性向上**
両サービス間でデータの不一致が解消される(ただしアプリ自体は別々)。
- **ユーザーによるデータ編集が可能**
ミスの修正をユーザーが提案でき、数日以内に反映。
- **データの微細な変更**
引用数や OA リンクなど、全体で約 5%のデータが変化。
- **新しいトピックアルゴリズム**
移行後に作成された研究成果は新アルゴリズムで分類(従来と同じ分類体系)。

- **新しいキーワード付与**
新しい概念アルゴリズムに基づくキーワードが追加。

変更されない点

- **ID の安定性**
OpenAlex の ID で取得できる対象は移行前後で不変。
- **機能性**
API、Web UI、スナップショット等の基本機能は維持。
- **データスキーマ**
データ構造自体は変更なし。

ロールアウトスケジュール

日付	内容
6月1日	Unpaywall 統合
10月1日	ベータ版公開 (data_version=2 パラメータで新データ利用可、比較ツール・ベータスナップ公開)
11月1日	本番公開 (新コードベースがデフォルト、旧データは data_version=1 で利用可能、プロダクションスナップ公開)
12月1日	完全移行 (旧データ API 提供終了、比較ツール終了、旧データの最終スナップ公開)

データ変化例

- 一部論文の引用数増減
- OA リンクの新規追加
- キーワード・トピックのアルゴリズム変更

最新情報の取得

- 開発は活発に進行中で、予期せぬ変更も発生する可能性あり。
- 最新情報は [OpenAlex ユーザーGoogle グループ](#) で随時更新。

参考リンク

- [公式ブログ記事](#)
- [Microsoft Academic Graph 終了について](#)
- [Apache Spark 公式](#)
- [Databricks 公式](#)

著者・情報元

- 投稿者: Jason Priem