

TIB-arXiv: Ein alternatives Suchportal für den arXiv Preprint Server

Matthias Springstein¹, Huu Hung Nguyen¹, Anett Hoppe¹, und Ralph Ewerth^{1,2}

¹Leibniz-Informationszentrum Technik und Naturwissenschaften (TIB), Hannover

²L3S Forschungszentrum, Leibniz Universität Hannover, Hannover

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten

- Aktuelle und relevante Artikel sind immer schwerer zu finden
 - Anzahl der Publikationen steigt von Jahr zu Jahr an
 - Diverse, neue Publikationsforen entstehen
- Zusätzliche Ressourcen sind nicht verknüpft (Quelltext)
- Soziale Medien haben keinen Einfluss auf Relevanz

Motivation

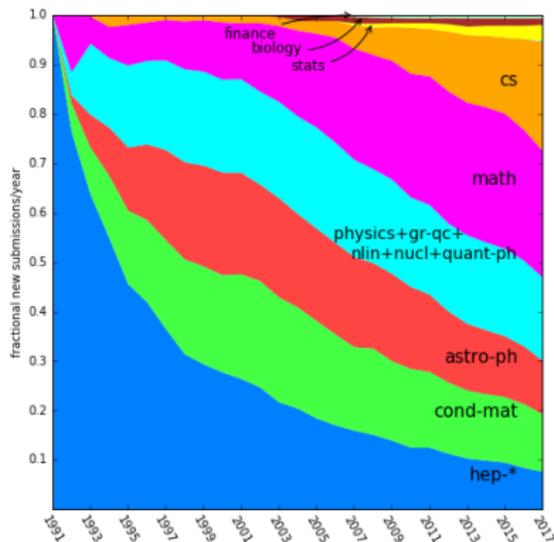
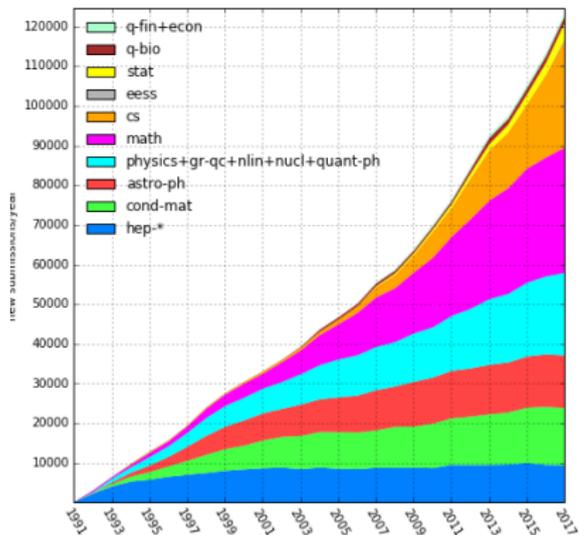


Abbildung: Anzahl der Einreichungen auf *arXiv* pro Jahr¹

¹https://arxiv.org/stats/monthly_submissions

- Bessere Web-Oberfläche zur Suche auf der *arXiv*-Plattform
- Verlinkung der Artikel zu externen Ressourcen
- Volltext-Suche für alle verfügbaren Artikel
- Verschiedene Ranking-Möglichkeiten

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten

Arxiv Sanity Preserver²

- Fokus auf Computer Vision und maschinellem Lernen
- Verschiedene Such- und Sortiermöglichkeiten

Arxivsorter³

- Fokus auf Artikel der Astrophysik
- Darstellung von einzelnen Abbildungen

²<http://www.arxiv-sanity.com>

³<http://www.arxivsorter.org>

PaperScape⁴



- Darstellung der Artikel auf einer 2D-Karte
- Zitate bestimmen Größe und Position des Artikels

⁴<http://paperscape.org>

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten

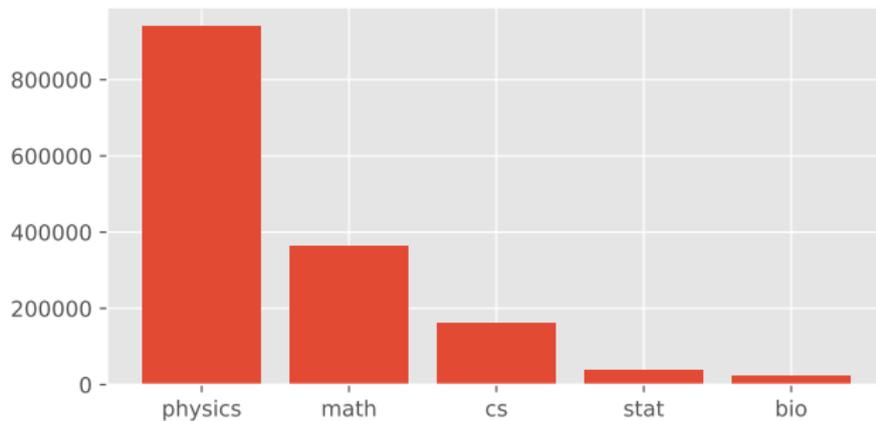


Abbildung: Verteilung der Artikel nach Fachbereich

- e-Print Server für MINT-Fächer
- 1,4 Mio. Artikel

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten

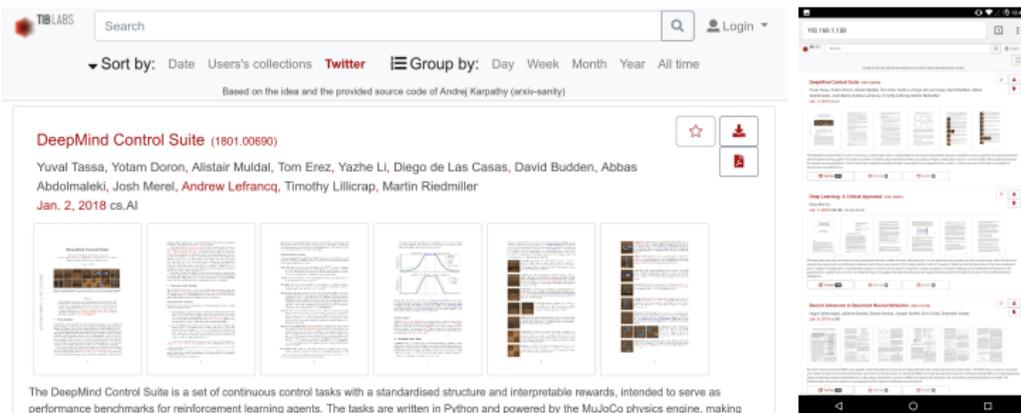
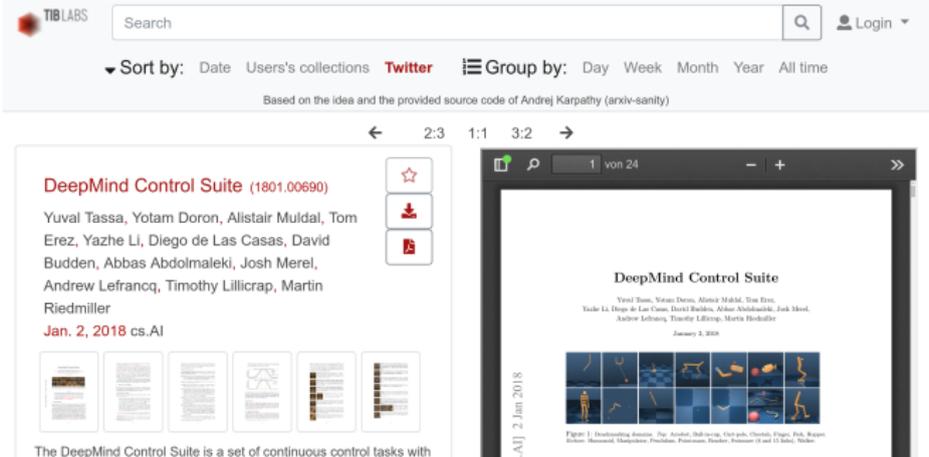


Abbildung: Benutzer-Interface Desktop und mobil

- Unterstützung von verschiedenen Auflösungen
- Einzelseiten-Anwendung



The screenshot displays the TIB arXiv PDF viewer interface. At the top, there is a search bar with the TIB LABS logo on the left and a 'Login' button on the right. Below the search bar, there are sorting options: 'Sort by: Date Users's collections **Twitter**' and 'Group by: Day Week Month Year All time'. A subtitle reads 'Based on the idea and the provided source code of Andrej Karpathy (arxiv-sanity)'. The main content area is split into two panels. The left panel shows search results for 'DeepMind Control Suite (1801.00690)', listing authors (Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew LeFrancq, Timothy Lillicrap, Martin Riedmiller) and the date 'Jan. 2, 2018 cs.AI'. It includes a grid of thumbnail images and icons for favorite, download, and share. The right panel shows a preview of the PDF document, which is titled 'DeepMind Control Suite' and features a grid of colorful images representing various control tasks. The PDF viewer interface includes navigation arrows and a page indicator '1 von 24'.

Abbildung: Suchergebnis mit zugehöriger PDF-Datei

- Mozilla PDF.js als PDF Betrachter
- Verschiedene Aufteilungen zwischen Ergebnisliste und PDF

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten

- Elasticsearch als Open-Source-Volltextsuche
- Indizierung aller Meta-Informationen von *arXiv*
- Einschränkung der Suche auf Suchfelder möglich

Sortierfunktion nach unterschiedlichen Kriterien:

- **Datum:** Erscheinungsdatum der letzten Version
- **Twitter:** Anzahl der Erwähnungen auf Twitter
- **Collection:** In wie vielen Benutzerbibliotheken ist ein Artikel
- **Relevanz:** Übereinstimmung mit Suchanfrage

1. Motivation und Zielsetzung
2. Verwandte Arbeiten
3. ArXiv
4. User Interface
5. Such- und Rankingfunktionen
6. Zukünftige Arbeiten



reddit



- Verlinkung der Artikel zu neuen externen Ressourcen
 - Relevante Repositorien auf GitHub
 - Diskussionen auf Reddit
- Import und Export von Mendeley

Künftig: Verbessertes Empfehlungssystem



- Empfehlung neuer Artikel anhand eigener Bibliothek
- Merkmalsextraktion:
 - Einfachste Lösung nutzt Tf-Idf
 - Deep Learning für bessere Merkmale
- Aktuelle Verfahren extrahieren mehr Kontext

Vielen Dank für Ihre
Aufmerksamkeit.



<https://labs.tib.eu/arxiv>