# Zentralblatt MATH – on the way

ÖMG-DMV - Tagung Graz, 21. September 2009

Olaf Teschke



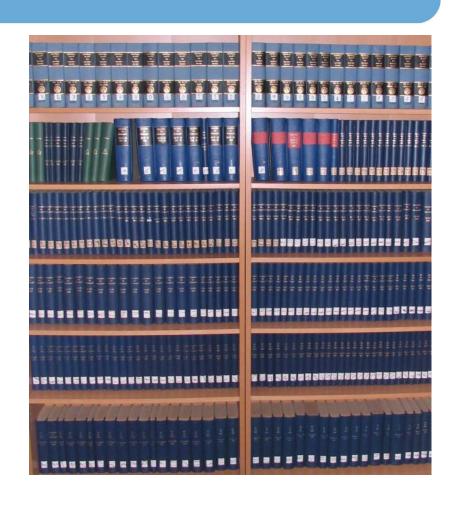


# Ein Jugendtraum

Wie kann ich

"alles"

über Mathematik wissen?





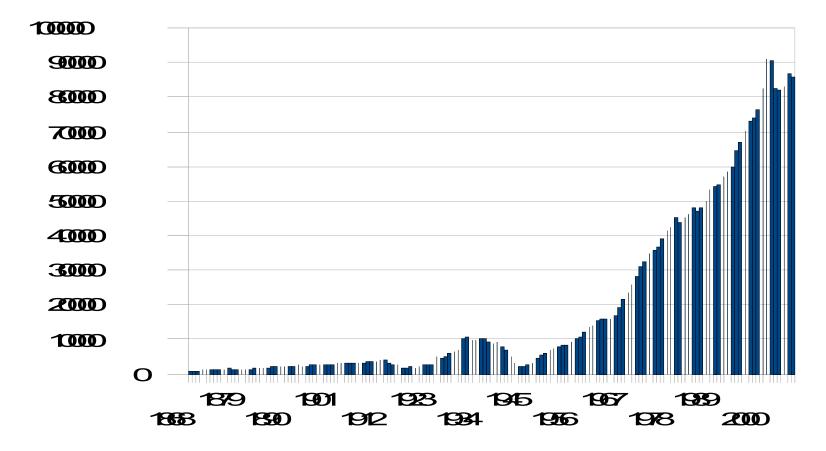


Ich lebe vor 100 Jahren und heiße Henri Poincaré.













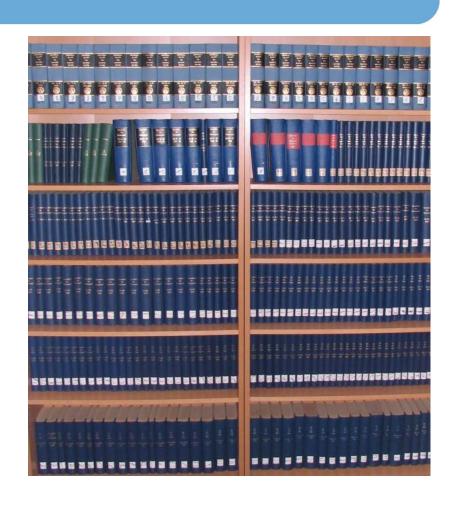
# Ein Jugendtraum (2)

Wie kann ich

"alles"

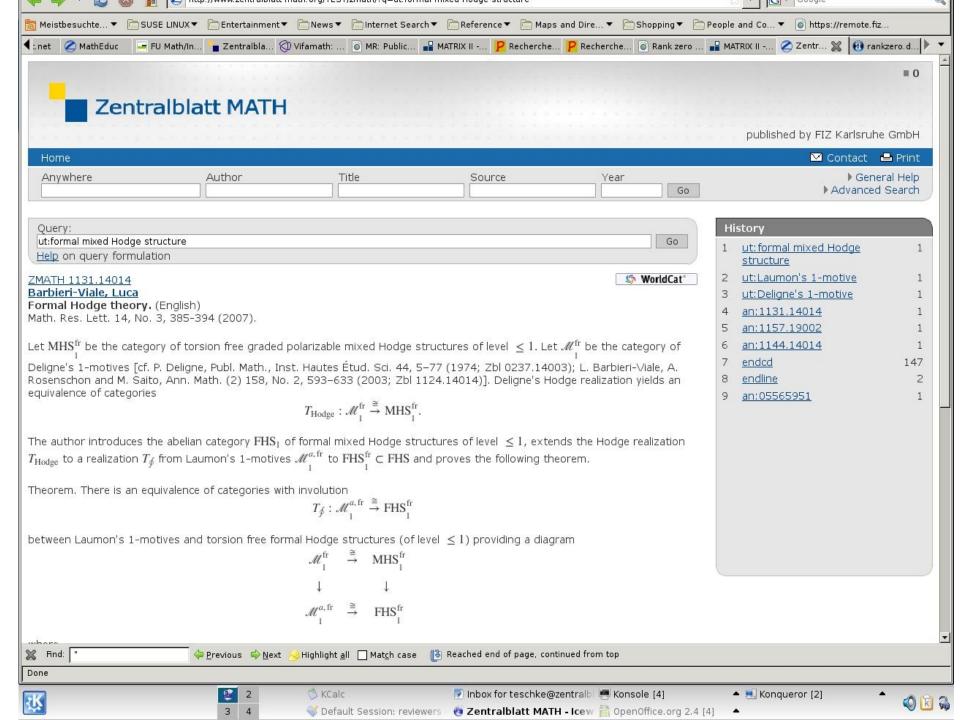
über Mathematik wissen,

was ich will?









#### Die Mühen der Ebene

In der harten Realität stellt schon die Aufbereitung scheinbar einfachster Informationen eine Herausforderung dar.

Die simple Frage "Wer ist Autor der Arbeit" führt auf komplexe Fragen der Autorenidentifikation (Schreibweisen, Transkription, Homonyme...)

| Au                               | thor Name:  | Author-ID:                 |                  |
|----------------------------------|---|----------------------------|------------------|
| re                               | cillas, sevin   | 1                          | 11               |
| Enter a query and click »Search« |   | Clear                      | Search           |
| Re                               | sult: 1 authors identified  |                            |                  |
|                                  | Recilias Pishmish, Sev\'{\i}n □ Autho<br>Count: 23 ]<br>Spellings: Recilias, Sevin; Recilias, S.; F<br>Recilias-Pishmish, Sevin; Recilias-Pishm | tecillas, Sev\'{\i}n; Reci | llas, Sev\'\i n; |





# Die Mühen der Ebene (2)

Lösbar sind diese Probleme heute nur durch die Kombination automatischer Verfahren mit intellektueller Aufbereitung. Automatisch können Schreibweisen nach Regeln zusammengeführt werder, ebenso hilft bei Zusammenführung /Trennung die Auswertung der Koautoren und der Arbeitsgebiete (durch die Klassifikation). Trotzdem kann im kritischen Einzelfall die Zuordnung nur intellektuell sicher erfolgen.





# Die Mühen der Ebene (3)

Nachdem mit Beginn 2008 nur (hochgradig) sicher identifizierte Autoren fest zugeordnet wurden, besteht in der 2. Version der Autorenidentifikation 2010 mit der Einführung verschiedener Identifikationsgrade die Möglichkeit der Komplettzuordnung. Schwach identifizierte Autorennamen sind dabei evtl. mehreren Identitäten zugeordnet.





### Integration und Vernetzung

Auf ähnliche Weise erfolgt auch die Zuordnung zu Journalen (besonders interessant auch hier die Aufbereitung historischer Quellen). Dies bildet die Grundlage von

- Verlinkung der Volltexte (digitale Bibliotheken)
- Anreicherung durch Referenzen und deren Aufbereitung (Generierung von Zbl-Stellen/DOI)
- Integrierte Suche





#### Kontrolliertes Vokabular

Keyword-Analyse und Aufbau eine kontrollierten Vokabulars:

Ausgangssituation:

Klassenbezeichner der MSC (ca 10.000 Terme) vorhandene Keywords in ZMATH (ca. 3.300.000 Terme)

Rohmaterial für ein kontrolliertes Vokabular,

Aufbereitung erforderlich

Software für die automatische Keyword-Extraktion (z.B. KEA),





# Mathematik ist (auch) Formel

LaTex ist heute (auch) Standardcode für die Digitalisierung mathematischer Formeln. Zur lesbaren Darstellung im Web empfiehlt sich aber die Umsetzung in MathML.

Bei der künftigen semantischen Erschließung liegt der nächste Schritt in der Anreicherung durch OpenMath+MathML.





## Mathematik ist (mehr und mehr) Algorithmus

Publikationen repräsentieren nur mehr einen Teil der Forschung. Algorithmen und Computerprogramme sind mehr als nur ein Beiwerk zur Publikation, sondern zunehmend eigenständige Forschungsergebnisse mit eigenen Lebenszyklen und Metadaten.

Die strukturierte Erfassung steht hier noch am Anfang, ebenso wie für automatische Beweise.





## Looking back forward

Für bedeutende historische Arbeiten lohnt sich auch die nachträgliche intellektuelle Aufbereitung (Klassifikation, second reviews).

In der Zukunft wird die Iteration der Referate durch Kommentarfunktionen zentrale Bedeutung gewinnen (``ZMATH 2.0").





#### Vielen Dank!



