KOPI Protection
Instead of Copy Protection

Máté Pataki

MTA SZTAKI
Topics

- Plagiarism
- KOPI Portal
- How KOPI Works
- KOPI Protection
- Future Plans
Problems

- Plagiarism is a huge problem at universities.
- There are too many theses even at one university, no one can be familiar with all of them.
- It is not enough to feel that something could be a plagiarism, some proof is needed.
Problems - Existing Systems

- Watermark or checksum
- Authorship attribution
- Open search engines
- Text comparison
- Questionnaire
- Systems with unknown algorithms
- No system for the Hungarian community
What we need

- Detects Partial Overlapping
- Can not be automatically removed
- Language independent
- Can protect proprietary documents
- One to many comparison
- Without user intervention
- Known algorithm
Topics

- Plagiarism
- KOPI Portal
- How KOPI Works
- KOPI Protection
- Future Plans
KOPI Online Plagiarism Search and Information Portal – Web based similarity and plagiarism search service

Partner: Monash University, Melbourne

Sponsored by the Hungarian Government

Developed 2003-2004

The Service is freely available to everybody
The Goal of KOPI

- Protect digital libraries from illegal copying
- Help teachers, professors, conference organizers to easily find copied work, and the original source
- Inform students and authors about plagiarism and citations and the relevant (Hungarian) laws
- Increase the values of papers, theses by certifying their genuineness
- Compare uploaded documents to each other
- Find similar documents on the database of the system:
  - Within the users own documents
  - Documents uploaded by others
  - Documents from the Internet
  - Digital libraries (MEK)
  - Universities
  - ...
Topics

- Plagiarism
- KOPI Portal
- How KOPI Works
- KOPI Protection
- Future Plans
How it works

1. **Chunking**
   - **text** → **chunk**

2. **Compress (MD5)**
   - **chunk** → **fingerprint**

3. **Upload to DB**
   - **fingerprint** → **DB**

4. **Query**
   - **DB** → **result**
The goal of the KOPI online Plagiarism Search and Information Portal is to protect documents against plagiarism.

Word chunking (n=5)

the goal of the kopi
online plagiarism search and information
portal is to protect documents
The goal of the KOPI online Plagiarism Search and Information Portal is to protect documents against plagiarism.

Overlapping word chunking (n=5)

- the goal of the kopi
- goal of the kopi online
- of the kopi online plagiarism
- the kopi online plagiarism search
- kopi online plagiarism search portal
- …
Compressing fingerprints

Hash based algorithm (MD5)

- Input length is not limited
- Fast
- The chance of two different texts to have the same MD5 code is small
- Irreversible
- Can protect proprietary documents
Topics

- Plagiarism
- KOPI Portal
- How KOPI Works
- KOPI Protection
- Future Plans
Copy protection

**Pros**
- Harder to copy it
- The way of the work can be followed (DRM)
- More income for authors and sellers

**Cons**
- Harder to use it
- Can not totally prevent copying
- Sometimes for the legal use it must be circumvented
- It is not always legal to use
- Personal rights problems (DRM)
- Hinders the spreading of the work
Copy Protection for Text Documents

- PDF, DOC… protection
  - Can be easily and automatically circumvented
- Allow only online viewing
  - Strongly restricts the use
  - It is harder, but can be circumvented
- Narrow down the number of authorized users
  - If once the documents is out of the system…
- Nothing protects against typing down
  - Close up into a drawer and leave it there
KOPI Protection

- Documents uploaded into the KOPI System
  - Plagiarism can be easily discovered
  - The sources will also be known
  - The risk to plagiarize will be too high
  - Circumventing it is time consuming and can not be done automatically

- The work can be freely distributed
  - Must not deal with copy protection
  - Search engines can index it
  - More people read it
  - More people cite from it
Topics

- Plagiarism
- KOPI Portal
- How KOPI Works
- KOPI Protection
- Future Plans
Future Plans

- Distributed System
  - Each university has its own system, but
  - They are able to search in the others' DB
  - Secure search with MD5 codes
- Upload databases
  - Online and offline databases
  - Documents found on the Internet
- Recognizing source codes and programming languages
- SOAP interface for integrated use of KOPI
Future Plans

- Portal Engine
  - Forum
  - FAQ
  - Help
  - User MGM.

- Plagiarism Search Engine

- Message Handler

- Document Upload and Management

- Document Interoperability

- Document Repository

- Document Converter

- Other KOPI Servers
http://kopi.sztaki.hu
Thank you for your attention!

Web: http://dsd.sztaki.hu

Email: Mate.Pataki@sztaki.hu