

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

FbHash: shema za izračun podobnosti datotek v digitalni forenziki

Timotej Knez¹, Sebastian Mežnar², Jasmina Pegan¹

¹Fakulteta za računalništvo in informatiko
Univerza v Ljubljana

²Fakulteta za matematiko in fiziko
Univerza v Ljubljana

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ Uvod
- ▶ Sorodna dela
- ▶ Algoritem
- ▶ Eksperimenti v članku
- ▶ Naši eksperimenti
- ▶ Zaključek

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ Avtomatizacija preiskave datotek
- ▶ Algoritmi za iskanje približnega ujemanja
 - ▶ ssdeep, sdhash, FbHash
- ▶ Prispevki članka
 - ▶ odporna shema, dve različici, analiza varnosti
- ▶ Implementacija algoritma
 - ▶ različne funkcije za uteževanje
- ▶ Testiranje na istih množicah kot v članku

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ ssdeep
 - ▶ Temelji na algoritmu spamsum
 - ▶ Funkcija z drsečim oknom
- ▶ sdhash
 - ▶ Statistično najmanj verjetni deli datoteke
 - ▶ Bloomovi filtri
- ▶ MRSH-v2
 - ▶ Multi-resolution similarity hashing
 - ▶ Funkcija z drsečim oknom, Bloomovi filtri
- ▶ mvHash-B
 - ▶ Krajšanje zapisa: glasovanje, kompakten zapis
 - ▶ Bloomovi filtri

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ FbHash-B za nestisnjene in FbHash-S za stisnjene datoteke
- ▶ Glavna komponenta so deli, sestavljeni iz k zlogov

$$\begin{aligned}
 \text{ch}_0^D &= B_0^D, B_1^D, B_2^D, \dots, B_{k-2}^D, B_{k-1}^D \\
 \text{ch}_1^D &= B_1^D, B_2^D, B_3^D, \dots, B_{k-1}^D, B_k^D \\
 \text{ch}_2^D &= B_2^D, B_3^D, B_4^D, \dots, B_k^D, B_{k+1}^D \\
 &\vdots \\
 \text{ch}_i^D &= B_i^D, B_{i+1}^D, B_{i+2}^D, \dots, B_{i+k-2}^D, B_{i+k-1}^D \\
 &\vdots
 \end{aligned}$$

Figure: Oblika delov datoteke

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

Algoritem FbHash-B:

- 1 Za vsak del datoteke izračunaj zgoščeno vrednost in preštej pojavitve.
- 2 Uravnoteži dele datoteke glede število pojavitev
- 3 Preštej in uravnoteži pojavitve delov v bazi podatkov
- 4 Izračunaj končno oceno za datoteko iz uteži

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

► Funkcije za uteži delov znotraj datoteke:

$$\text{► } W_{ch_i}^D(chf_{ch}^D) = 1 + \log_{10} \left(\frac{chf_{ch}^D}{n} \right)$$

$$\text{► } W_{ch_i}^D(chf_{ch}^D) = \log_2 \left(1 + \frac{chf_{ch_i}^D}{n} \right)$$

$$\text{► } W_{ch_i}^D(chf_{ch}^D) = \frac{chf_{ch_i}^D}{n}$$

► Funkcije za uteži delov v podatkovni bazi

$$\text{► } docw_{ch_i}^D(df_{ch}) = \log_{10} \left(\frac{m}{df_{ch}} \right)$$

$$\text{► } docw_{ch_i}^D(df_{ch}) = 1 - \log_{10} \left(\frac{df_{ch}}{m} \right)$$

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
žlanku

Naši
eksperimenti

Zaključek

- ▶ Primerjava datotek s kosinusno razdaljo

$$\textit{Similarity}(D_1, D_2) = \frac{\sum_{i=0}^{n-1} W_{ch_i}^{D_1} \cdot W_{ch_i}^{D_2}}{\sqrt{\sum_{i=0}^{n-1} W_{ch_i}^{D_1}} \cdot \sqrt{\sum_{i=0}^{n-1} W_{ch_i}^{D_2}}} \cdot 100$$

- ▶ FbHash-S uporabi FbHash-B na razširjenih datotekah

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ Algoritem varen proti napadu z aktivnim napadalcem
- ▶ Detekcija fragmentov
- ▶ Korelacija skupnega dela datoteke

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

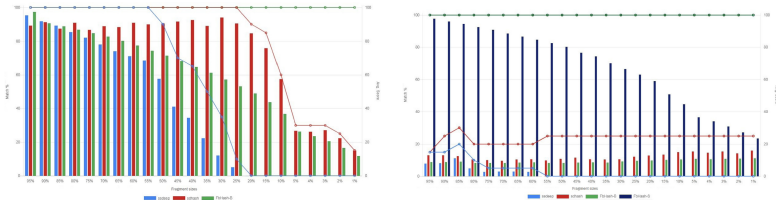


Figure: Rezultati detekcije fragmentov iz članka

Korelacija skupnega dela datoteke

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

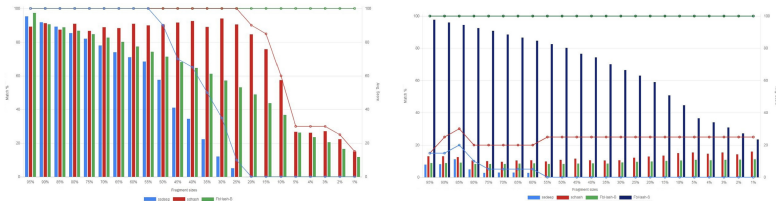


Figure: Rezultati korelacije skupnega dela datoteke iz članka

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ Implementacija algoritma v jeziku python
- ▶ Implementacija fbhash-s in fbhash-b
- ▶ Testiranje na prosti zbirki dokumentov t5 (<http://roussev.net/t5/t5.html>)
- ▶ Testiranje delovanja algoritma pri različnih funkcijah za izračun uteži

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Ekspirimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ Generiranje datotek z znanim ujemanjem
 - ▶ Vzamemo dve datoteki iz zbirke
 - ▶ Naključni blok prve vstavimo v drugo
 - ▶ Primerjamo prvo datoteko z novo nastalo datoteko
- ▶ Algoritem naučimo na celotni zbirki dokumentov
- ▶ Z algoritmom primerjamo ustvarjena dokumenta
- ▶ Opazujemo
 - ▶ Delež datotek z zaznanim ujemanjem
 - ▶ Povprečno oceno ujemanja glede na resnično ujemanje
 - ▶ F-oceno

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

**Naši
eksperimenti**

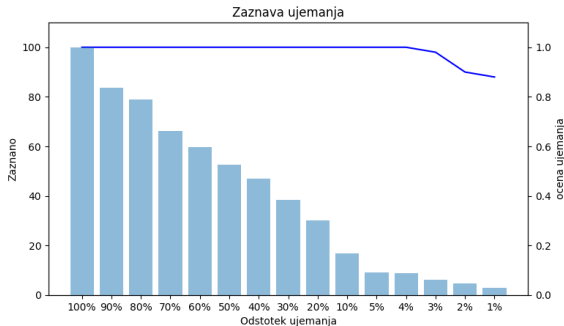
Zaključek

► Izračun uteži kot v izvornem članku

$$\text{► } W_{ch_i}^D(chf_{ch}^D) = 1 + \log_{10} \left(\frac{chf_{ch}^D}{n} \right)$$

► F ocena: 0.94

► Rezultati skladni s člankom



FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

**Naši
eksperimenti**

Zaključek

► Uporaba različnih uteži

► Levo: $W_{ch_i}^D(chf_{ch}^D) = \log_2 \left(1 + \frac{chf_{ch_i}^D}{n} \right)$

► Desno: $W_{ch_i}^D(chf_{ch}^D) = \frac{chf_{ch_i}^D}{n}$

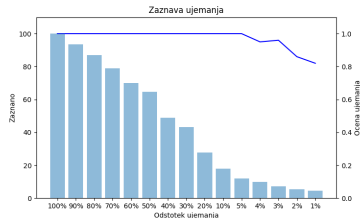
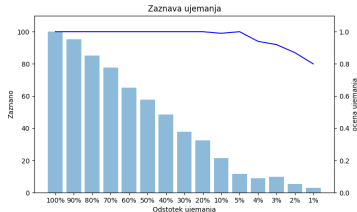


Figure: Testiranje delovanja algoritma z različnimi utežmi

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

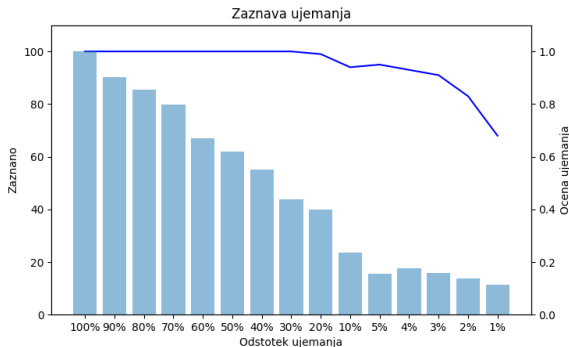
Algoritem

Eksperimenti v
članku

**Naši
eksperimenti**

Zaključek

- ▶ Razpoznavanje brez zbirke dokumentov
- ▶ F ocena: 0.87
- ▶ Dovolj dober rezultat, da bi lahko algoritem uporabili tudi brez zbirke dokumentov.



FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ The proposed approach improves the average and maximum relative errors compared to the existing square approximations.

FbHash

Knez, Mežnar,
Pegan

Kazalo

Uvod

Sorodna dela

Algoritem

Eksperimenti v
članku

Naši
eksperimenti

Zaključek

- ▶ The proposed approach improves the average and maximum relative errors compared to the existing square approximations.
- ▶ Error analysis has shown that an error in the circuit is directly proportional to the trouble it can cause.