

Реклами, реклами... – намиране без взирание

Мирослав Иванов¹, Красимира Иванова¹,
Илия Митов¹, Евгения Великова²

1: Институт по математика и информатика при БАН

2: Софийски университет "Св. Климент Охридски" - ФМИ

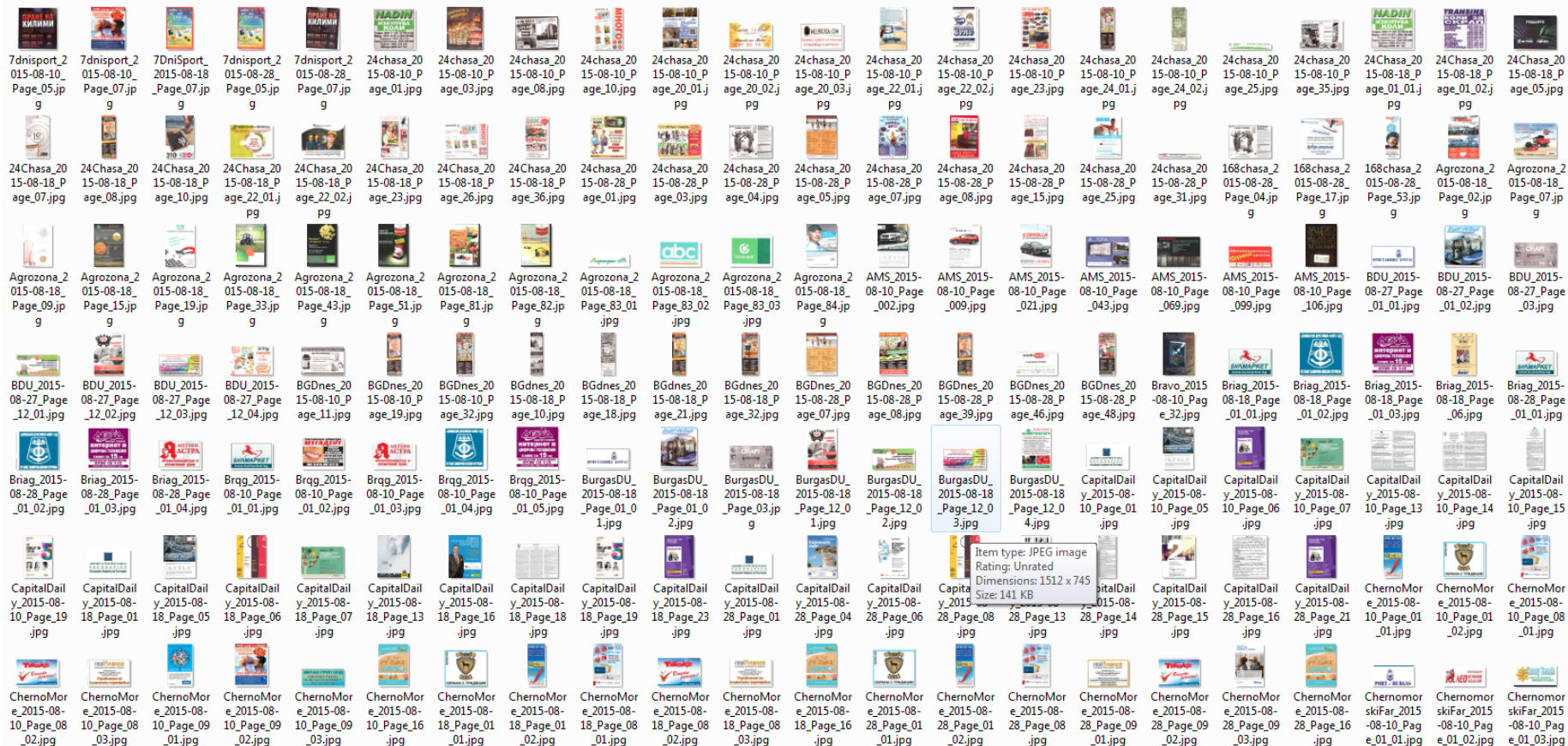
- Област
- Average Hash
- Приложимост ?
- Организация на базата

Медиен анализ

- Оценка на медийния отзвук на кампания или събитие
- Сравняване на медийното отразяване на две или повече компании
- Изчисляване на рекламната стойност на реклами и PR публикации

Бюлетин; реклами

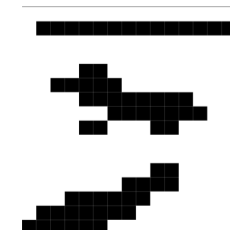
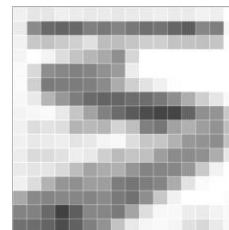
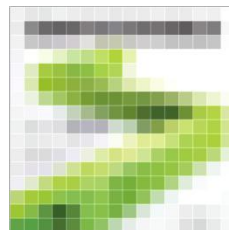
- Бюлетинът – съдържа:
 - обобщени количествени данни, оценки и анализи
 - преки референции към материалите, съдържащи името на клиента или следените конкуренти
- Печатните реклами:
 - Сканиране
 - "Нарязване"
 - Съхраняване в различни контейнери
 - Текстови
 - Изображения



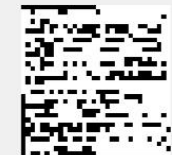
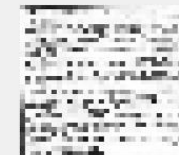
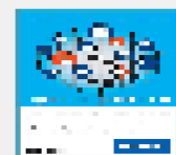
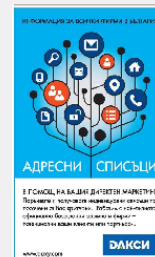
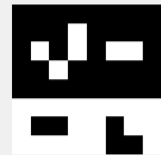
- В настоящата работа се предлага един полуавтоматизиран подход за организиране на процеса на преглеждане на изображенията за наличие на търсени реклами. Постъпващите изображения се сравняват с налични в базата етикетирани първообрази и при достатъчна близост те директно се категоризират в същата група. При срещане на изображение, което системата не може да причисли към някое от етикетираниите, се извършва ръчно категоризиране от оператора или отхвърлянето му като не-рекламен материал.

Average Hash

- Step 1 – Size reduce
- Step 2 – Color reduce (luminosity в YCbCr)
- Step 3 – Calculate Average Color
- Step 4 – Calculate the hash



pHash=(0000000000000000 0111111111111110 0000000000000000 0000000000000000
0000110000000000 00011111100000000 ... 0001111110000000 0111111100000000 1111110000000000)

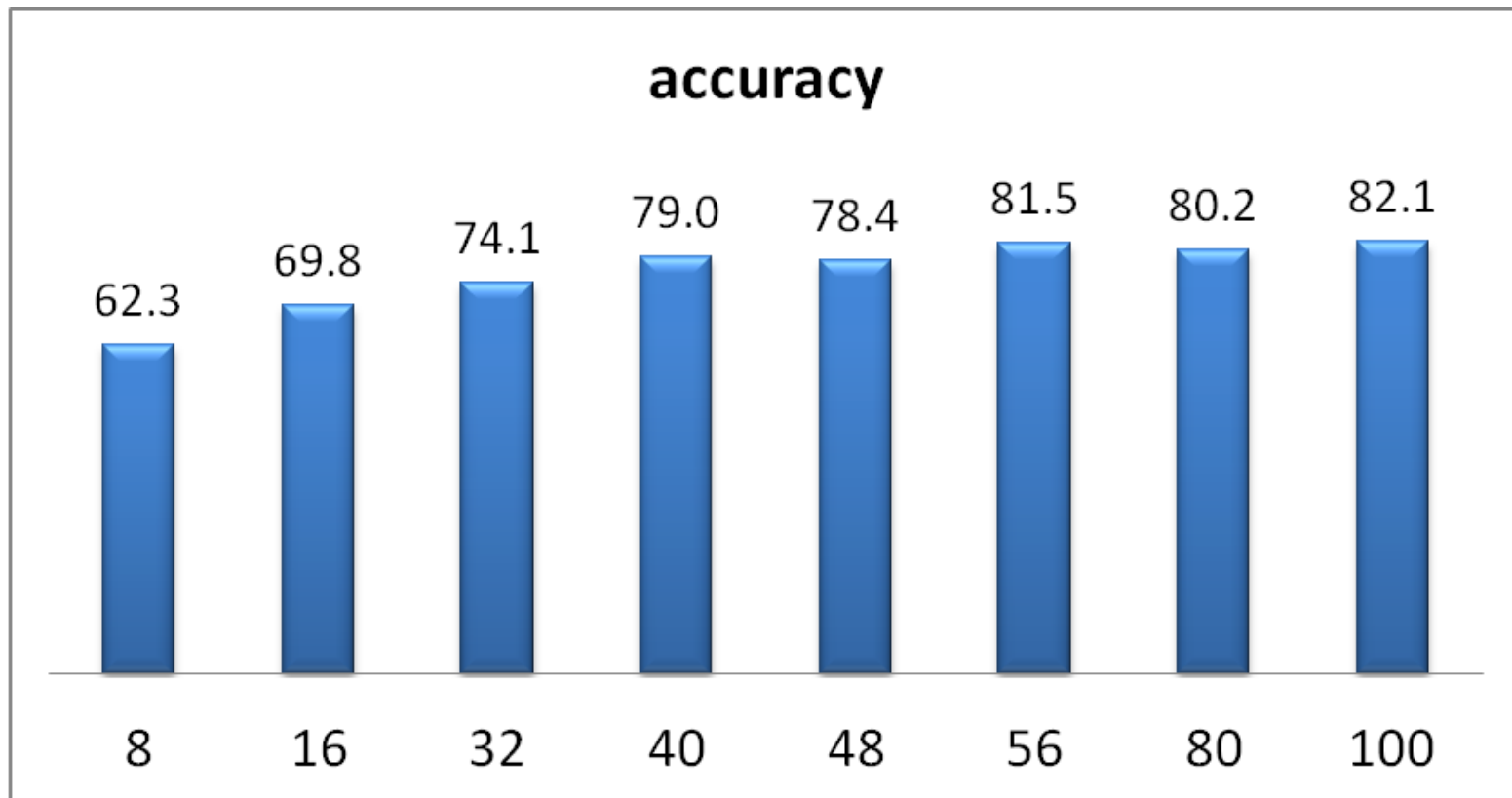


- Сравнението на две изображения става на базата на оценка на Хеминговото разстояние между двата описателя

$$d(p, q) = \sum_{i=1}^{l*h} (p_i - q_i)$$

Експеримент (1)

- Weka – IB1 – най-близък съсед!
- 98 – обучение; 162 – тестови. Всички са етикетирани.



Експеримент (2)

- Scheme:weka.classifiers.lazy.IB1
- Relation: hashfile-056
- Instances: 900
- Test mode:10-fold cross-validation
- === Summary ===
- Correctly Classified Instances 661 73.44 %
- Incorrectly Classified Instances 239 26.56 %

- Допълнителни:
 - обяви (50)
 - неетикетирани (590)

Организация на данните

- Първообрази:
 - Векторите, с които се сравняват новите изображения
 - Таговете, въведени от оператора при първото срещане
- Повторения:
 - Сочат към идентификаторите на съответните първообрази

Какво още

- Индексен механизъм по кода за оптимизиране на търсенето
- Оценка на разстоянието на покритие на кодовите думи
- Допълнително ускорение чрез търсене първо в същите издания, клъстера от издания и накрая в останалите
- Оценка на времевия интервал, в който средно се задържа дадена реклама с цел оптимизиране на търсенето