



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale  
2007-2013



# Platformă de e-learning și curriculum e-content pentru învățământul superior tehnic

**AEACD**

**19. Detectia elementelor constitutive ale documentelor:  
Spatii albe. Separatori. Linii. Cadre si tabele**

# Spatii albe

- Au un rol foarte important in detectia paragrafelor si a fontului
- Spatiile albe dintre caractere si cuvinte pot duce la detectia fontului caracterelor
- Tipuri de spatii albe:
  - Separatori de paragrafe;
  - Separatori de linii de text;
  - Separatori de cuvinte pentru detectia fontului

# Spatii albe

- Un algoritm folosit pentru detectia lor este bazat pe distante
- Harta de distanta reprezinta o colectie de valori ale distantelor pentru fiecare pixel
- Pentru un pixel negru valoarea distantei este 0, iar pentru un pixel alb valoarea distantei reprezinta numarul de pixeli albi intrecalati intre cei mai apropiati 2 pixeli negrii pe o directie ce contine pixelul alb de start

# Spatii albe

- Astfel se pot crea 3 tipuri de harti de distante:
  - Orizontala;
  - Verticala;
  - Pe o directie aleatoare;
- Aceste trei tipuri de harti se pot combina:
  - Foarte restrictiv – harta finala va fi minimul dintre hartiile alese;
  - Putin restrictiv – harta finala va fi maximul dintre hartiile alese;
  - Avansat – harta finala reprezinta maximul dintre combinatia liniara cu factori aleatori sau alesi special ale hartiilor constitutive;

# Spatii albe

- Harta finala este folosita pentru detectia fragmentelor principale din pagina
- Pornind de la aceste fragmente si folosind un prag ce se va mica progresiv se vor determina si sub-fragmentele acestora
- Algoritmul se termina in momentul in care cel mai mic fragment dorit este gasit
- Acest algoritm poate fi imbunatatit prin considerarea directiilor verticale si orizontale cat si inclinariile acestora cu + si - 5 grade

# Spatii albe

- Rezultatul algoritmului:

## CDU beschwört „Jobmaschine“

Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Von Rolf Dietrich Schwartz  
Die Bundesregierung sagt voraus, dass es 1998 ein Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

BONN, 11. März. Die Bundesregierung meinte in ihrem Jahreswirtschaftsaussichtswort, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

Wirtschaftsminister Dieter Borchardt (SPD) prophezeierte in seinem Jahreswirtschaftsaussichtswort, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

## Türkische Polizisten ohne Strafe

Gerecht in Manisa nennt Vorwurf der Folter nicht bewiesen

Von Gunter Kühn  
BONN, 11. März. Das Verbrechen in der westtürkischen Stadt Manisa hat die Mehrheit zum Polizisten von dem Vorwurf freigesprochen, von bis zu 15 Personen gefoltert. Die Polizei hat keine Beweise für die Folter gebracht. Einem der Angeklagten gegenüber, der sich weigerte, Aussagen zu machen, sagte Borchardt von einem „Schick“ und klagte die Regierung gegen die Türkei an. Er muss sich um einen Justizminister handeln.

Die Schüler waren im Dezember 1995 unter dem Vorwurf der Mitgliedschaft in einer linksgerichteten kommunistischen Organisation von der Polizei in Manisa festgenommen worden. Die Polizei hat keine Beweise für die Folter gebracht. Einem der Angeklagten gegenüber, der sich weigerte, Aussagen zu machen, sagte Borchardt von einem „Schick“ und klagte die Regierung gegen die Türkei an. Er muss sich um einen Justizminister handeln.

Die Türkei ist international bereits mehrfach wegen Folter kritisiert worden. Erst in der vergangenen Woche war ein neuer Fall in Manisa bekanntgeworden. Ein Mann hatte nach schweren Misshandlungen durch die Polizei einen Selbstmord begangen. Die Polizei hat keine Beweise für die Folter gebracht. Einem der Angeklagten gegenüber, der sich weigerte, Aussagen zu machen, sagte Borchardt von einem „Schick“ und klagte die Regierung gegen die Türkei an. Er muss sich um einen Justizminister handeln.

11,5 Prozent. Danach meinte Borchardt: „Das Szenario liegt heute im Bereich der Unsicherheit.“ Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

## Öffentlicher Dienst

Schlichter zeigen sich „gemäßig optimistisch“

BONN, 11. März. Die beiden Vorsitzenden der Schlichtungsstellen für den öffentlichen Dienst sind nach eigenen Angaben „gemäßig optimistisch“, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

## Öffentlicher Dienst

Schlichter zeigen sich „gemäßig optimistisch“

BONN, 11. März. Die beiden Vorsitzenden der Schlichtungsstellen für den öffentlichen Dienst sind nach eigenen Angaben „gemäßig optimistisch“, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

## CDU beschwört „Jobmaschine“

Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Von Rolf Dietrich Schwartz  
Die Bundesregierung sagt voraus, dass es 1998 ein Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

BONN, 11. März. Die Bundesregierung meinte in ihrem Jahreswirtschaftsaussichtswort, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

Wirtschaftsminister Dieter Borchardt (SPD) prophezeierte in seinem Jahreswirtschaftsaussichtswort, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

## Türkische Polizisten ohne Strafe

Gerecht in Manisa nennt Vorwurf der Folter nicht bewiesen

Von Gunter Kühn  
BONN, 11. März. Das Verbrechen in der westtürkischen Stadt Manisa hat die Mehrheit zum Polizisten von dem Vorwurf freigesprochen, von bis zu 15 Personen gefoltert. Die Polizei hat keine Beweise für die Folter gebracht. Einem der Angeklagten gegenüber, der sich weigerte, Aussagen zu machen, sagte Borchardt von einem „Schick“ und klagte die Regierung gegen die Türkei an. Er muss sich um einen Justizminister handeln.

Die Schüler waren im Dezember 1995 unter dem Vorwurf der Mitgliedschaft in einer linksgerichteten kommunistischen Organisation von der Polizei in Manisa festgenommen worden. Die Polizei hat keine Beweise für die Folter gebracht. Einem der Angeklagten gegenüber, der sich weigerte, Aussagen zu machen, sagte Borchardt von einem „Schick“ und klagte die Regierung gegen die Türkei an. Er muss sich um einen Justizminister handeln.

Die Türkei ist international bereits mehrfach wegen Folter kritisiert worden. Erst in der vergangenen Woche war ein neuer Fall in Manisa bekanntgeworden. Ein Mann hatte nach schweren Misshandlungen durch die Polizei einen Selbstmord begangen. Die Polizei hat keine Beweise für die Folter gebracht. Einem der Angeklagten gegenüber, der sich weigerte, Aussagen zu machen, sagte Borchardt von einem „Schick“ und klagte die Regierung gegen die Türkei an. Er muss sich um einen Justizminister handeln.

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

## Öffentlicher Dienst

Schlichter zeigen sich „gemäßig optimistisch“

BONN, 11. März. Die beiden Vorsitzenden der Schlichtungsstellen für den öffentlichen Dienst sind nach eigenen Angaben „gemäßig optimistisch“, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

## Öffentlicher Dienst

Schlichter zeigen sich „gemäßig optimistisch“

BONN, 11. März. Die beiden Vorsitzenden der Schlichtungsstellen für den öffentlichen Dienst sind nach eigenen Angaben „gemäßig optimistisch“, dass die im nächsten Jahr die Wirtschaftswachstum von bis zu drei Prozent und ein Jahresmittel von vier Arbeitsplätzen pro Kopf der Bevölkerung geben wird. Das kann mit dem Mittelweg verhandelt. Die Opposition sprach von „Schönfärberei“.

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

Die Bundesregierung prophezeit 200 000 neue Stellen / Opposition: Schönfärberei

# Spatii albe

- Un alt tip de algoritm baza pe distante functioneaza in felul urmator:
- Se calculeaza pentru fiecare pixel alb numarul de pixeli albi pana la primul pixel negru in sus, jos, stanga si dreapta
- Aceste valori sunt adaugate intr-o matrice si pe baza acestora se calculeaza trei praguri:

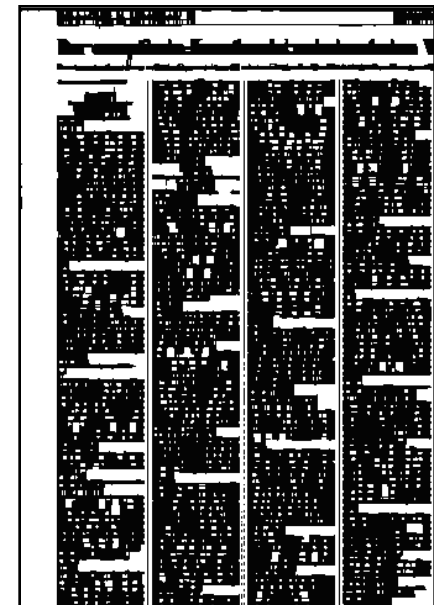
$$h = 1 - \frac{\textit{distanta\_sus} + \textit{distanta\_jos}}{\textit{inaltime} + \textit{latime}}$$

$$w = 1 - \frac{\textit{distanta\_stanga} + \textit{distanta\_dreapta}}{\textit{inaltime} + \textit{latime}}$$

$$\textit{suma} = w + h$$

# Spatii albe

- Inaltimea si latimea se refera la dimensiunile imaginii
- Pe baza acestor praguri, pixelii negri din imagine sunt uniti
- Mai jos se poate observa imaginea initiala impreuna cu rezultatul obtinut



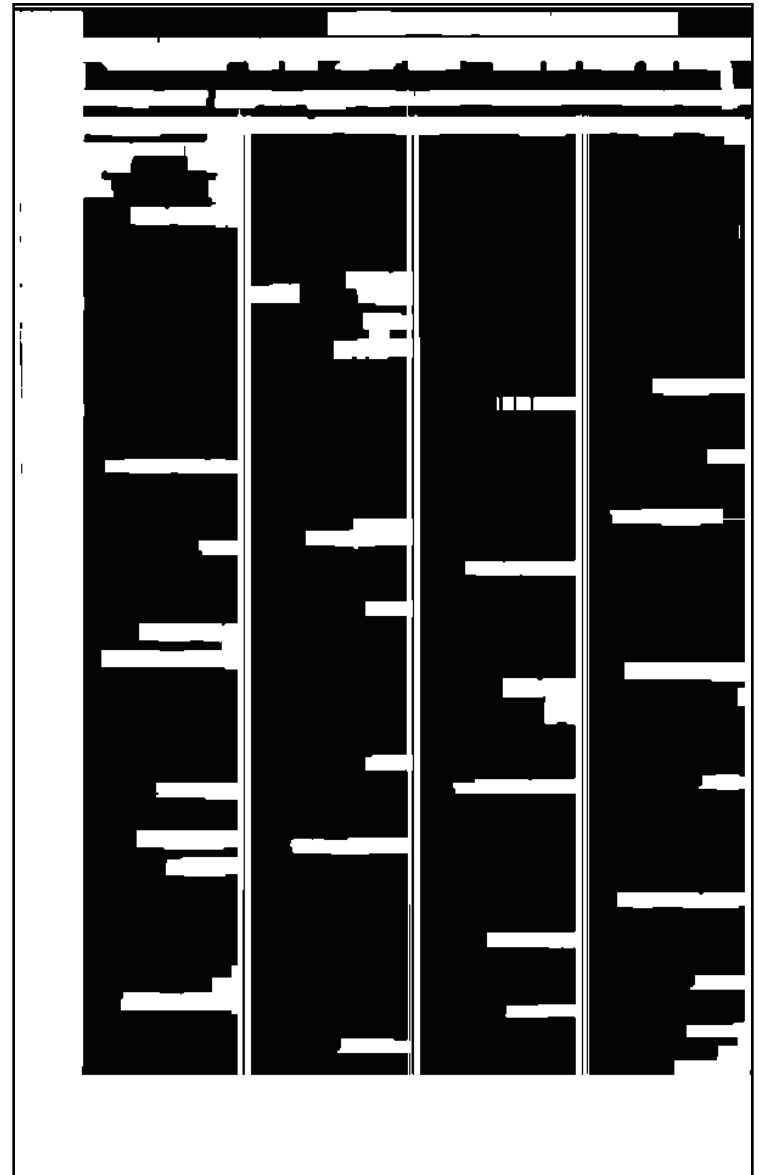


# Spatii albe

- Inca se pot observa mici gauri albe in interiorul paragrafelor ce trebuie eliminate
- Din acest motiv, se calculeaza numarul de treceri de la alb la negru pe toate cele 4 directii pentru fiecare pixel alb
- Pixelii din interiorul paragrafelor sunt impresmuiti de pixeli negrii deci numarul treceriilor va fi par
- Eliminand si aceste mici gauri din paragrafe, rezultatul final este obtinut si se poate aplica un simplu algoritm de detectie de contur pentru gasirea coordonatelor paragrafelor

# Spatii albe

- Rezultatul final:

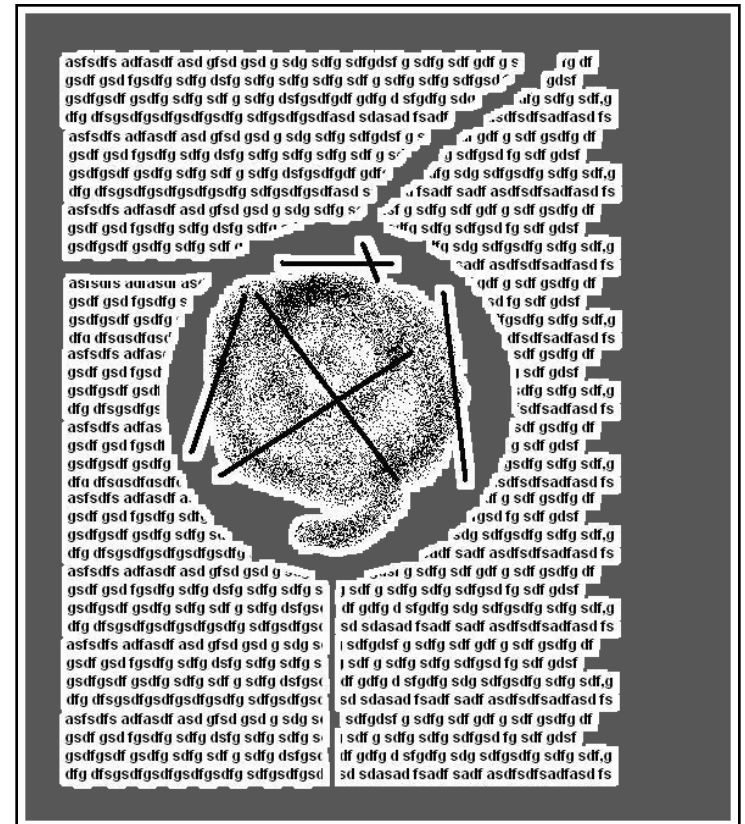
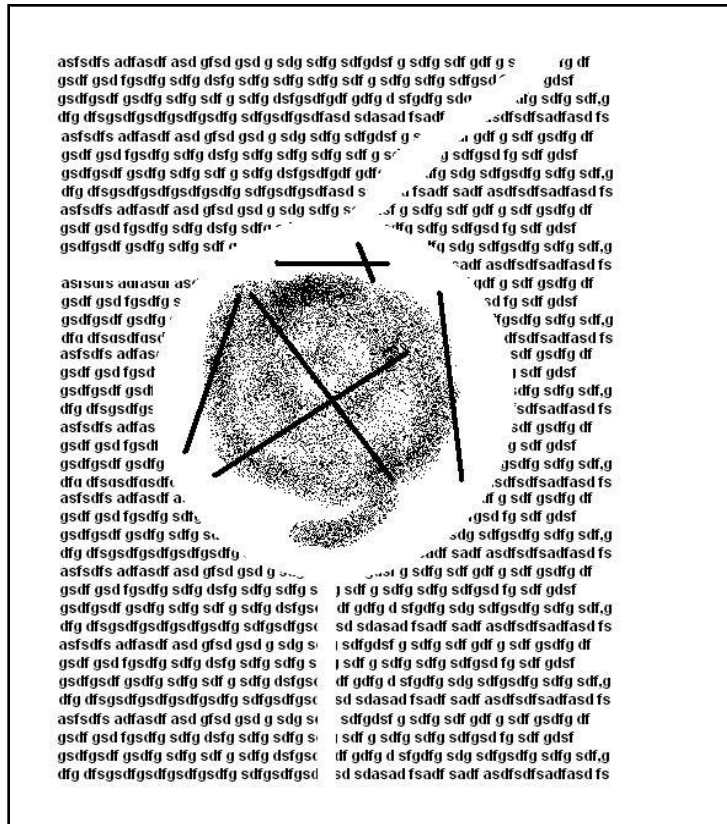


# Spatii albe

- Un alt tip de algoritm foloseste o metoda asemanatoare, insa ia in calcul si o masura ce se presupune deja cunoscuta marimea fontului
- Astfel se parcurge imaginea de la stanga la dreapta linie cu linie folosind un patrat ce are ca latura marimea fontului
- De fiecare data cand in interiorul patratului se vor gasi numai pixeli albi, vom putea spune ca acel patrat face parte dintr-un spatiu alb separator

# Spatii albe

- Spatiile gasite altfel vor fi separatori de paragraf si de linii de text



# Separatori

- Pot fi de mai multe tipuri
- Sunt foarte importanti pentru detectia layoutului
- Vor fi prezentati pe larg in capitolul urmator

# Linii

- Separatorii de tip linie vor fi prezentati pe larg in capitolul urmator
- Liniile de text se detecteaza verificand in parte fiecare doua entitati daca se afla pe acelasi rand
- Acest lucru se face verificand daca centrele fiecareia dintre cele doua entitati se afla intre limitele verticale ale celeilalte
- In caz afirmativ se adauga in vectorul de rand ce contine entitatile deja descoperite

# Linii

- Aceasta metoda ajuta la descoperirea spatiului dintre liniile de text (line spacing)
- Odata liniile detectate, caracterle ce se afla unul deasupra celuilalt ca pozitionare pe axa horizontala se verifica ca pozitionare pe axa verticala
- Daca distanta dintre ele este mai mica decat media inaltimeilor lor, atunci se afla pe linii de text consecutive si dimensiunea dintre linii este reprezentata de media distantelor dintre marginile inferioare ale caracterelor (base line)

# Cadre si Tabele

- Aceste se detecteaza folosind informatia deja existenta: linii, spatii albe
- O metoda generala pentru detectia de linii si cadre va fi prezentata in capitolul urmator

