

التعرف الضوئي على الكتابة العربية

شكرا لكل من : ماجد الرقيعي وعلاء الوهبي
وعبد اللطيف المهديب

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

الحوسبة العربية (تعريب الحاسبات) – ٢٣

التعرف الضوئي على الكتابة العربية المطبوعة

حسني المحتسب

التعرف الضوئي على الكتابة العربية

شكرا لكل من : ماجد الرقيعي وعلاء الوهبي
وعبد اللطيف المهديب

المسار

• مقدمة

• خطوات التعرف الضوئي

• مقارنة

• خاتمة

مقدمة

تحويل النص من الوثائق الورقيه الى ملف
حاسوب قابل للتعديل

تحويل النص من الوثائق الورقيه
الى ملف حاسوب قابل للتعديل

- ما هو التعرف الضوئي؟
- تحويل النص من الوثائق
الورقيه الى ملف حاسوب
قابل للتعديل
- أنواع التعرف الضوئي
- التعرف على الخط الثابت
- التعرف على خط اليد
- التعرف المباشر أو المتزامن

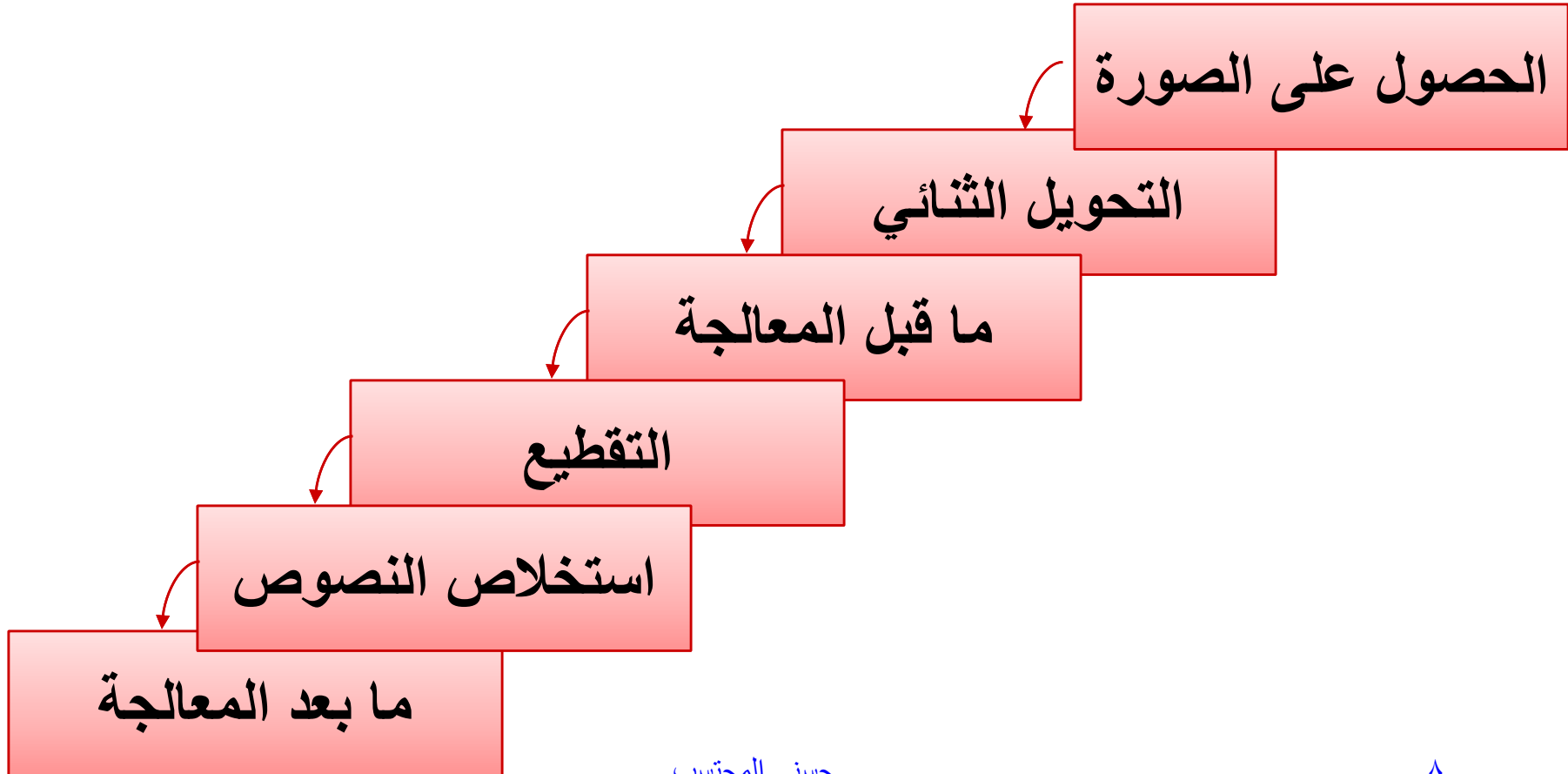
مقدمة

- التعرف الضوئي على الكتابه العربية (الخصائص)
 - الكتابة بشكل متصل
 - خطوط مختلفة
 - الأجسام الخارجية
 - الحروف تختلف على حسب موقعها
- أنواع برامج التعرف الضوئي
 - لا يحتاج للتدريب (OMNI)
 - يحتاج للتدريب

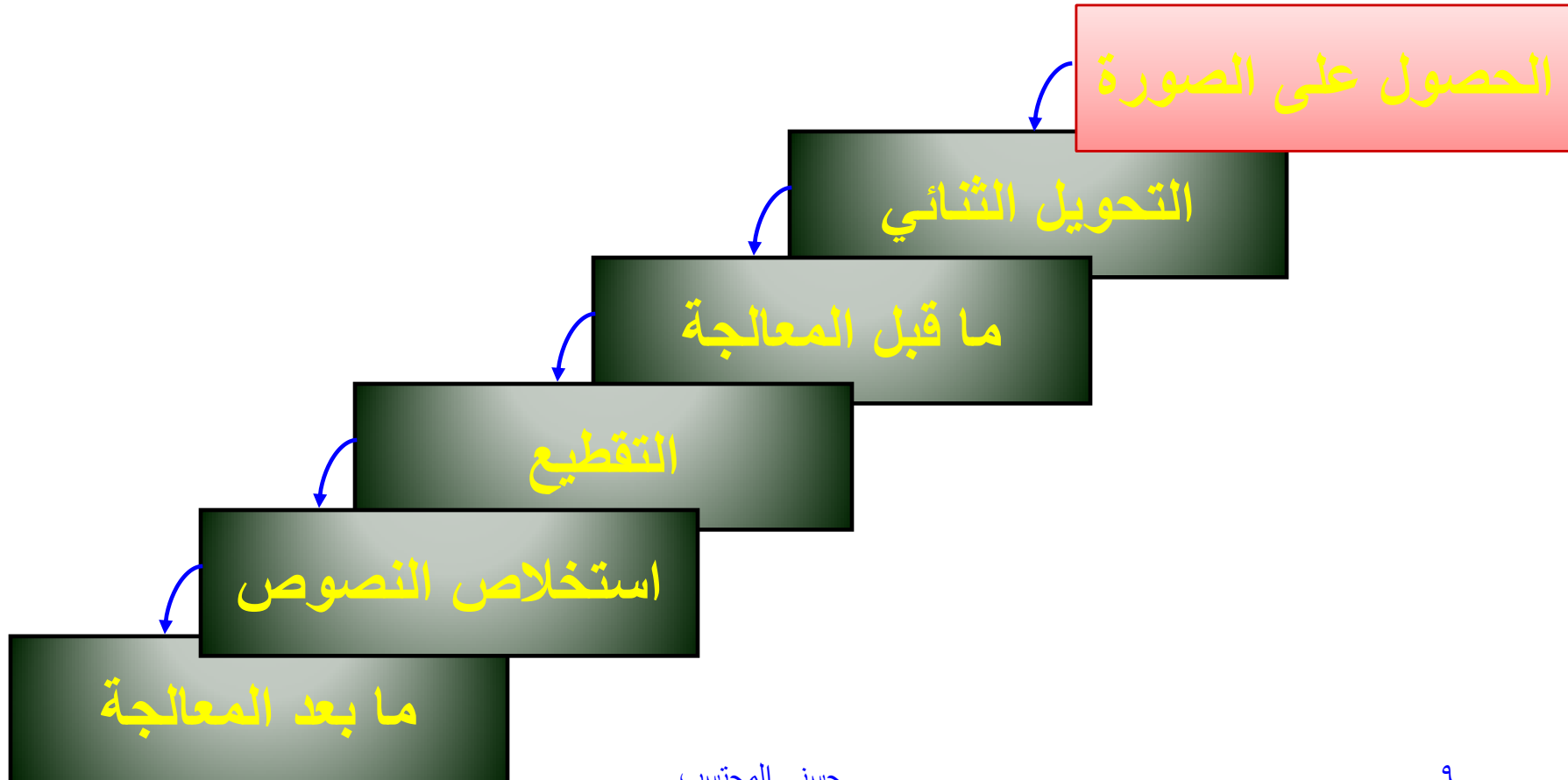
مقدمة

- ما أهمية التعرف الضوئي؟
- تحسين فعالية الحاسب للمستخدم
- تطبيقات المكاتب
- إدخال معلومات

التعرف الضوئي على الكتابة العربية – طريقة العمل



التعرف الضوئي على الكتابة العربية – طريقة العمل



التحويل الثنائي

• ما هو التحويل الثنائي؟

- غالبا ما تتم الكتابة على ورق أبيض وبخط أسود فتتم في هذه الحالة تحويل المناطق السوداء الى نقاط صغيرة تتمثل بالرقم ١ بينما يتم ترميز المناطق الفاتحة بالرقم ٠

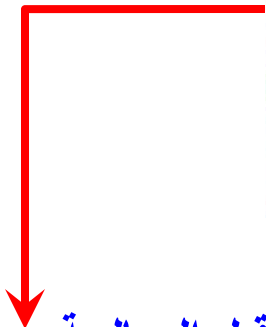
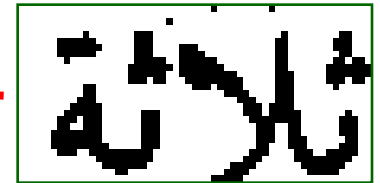
• لماذا تتم عملية التحويل الثنائي؟

- الأرقام الثنائية أسهل في التعامل من الصورة

التحويل الثنائي

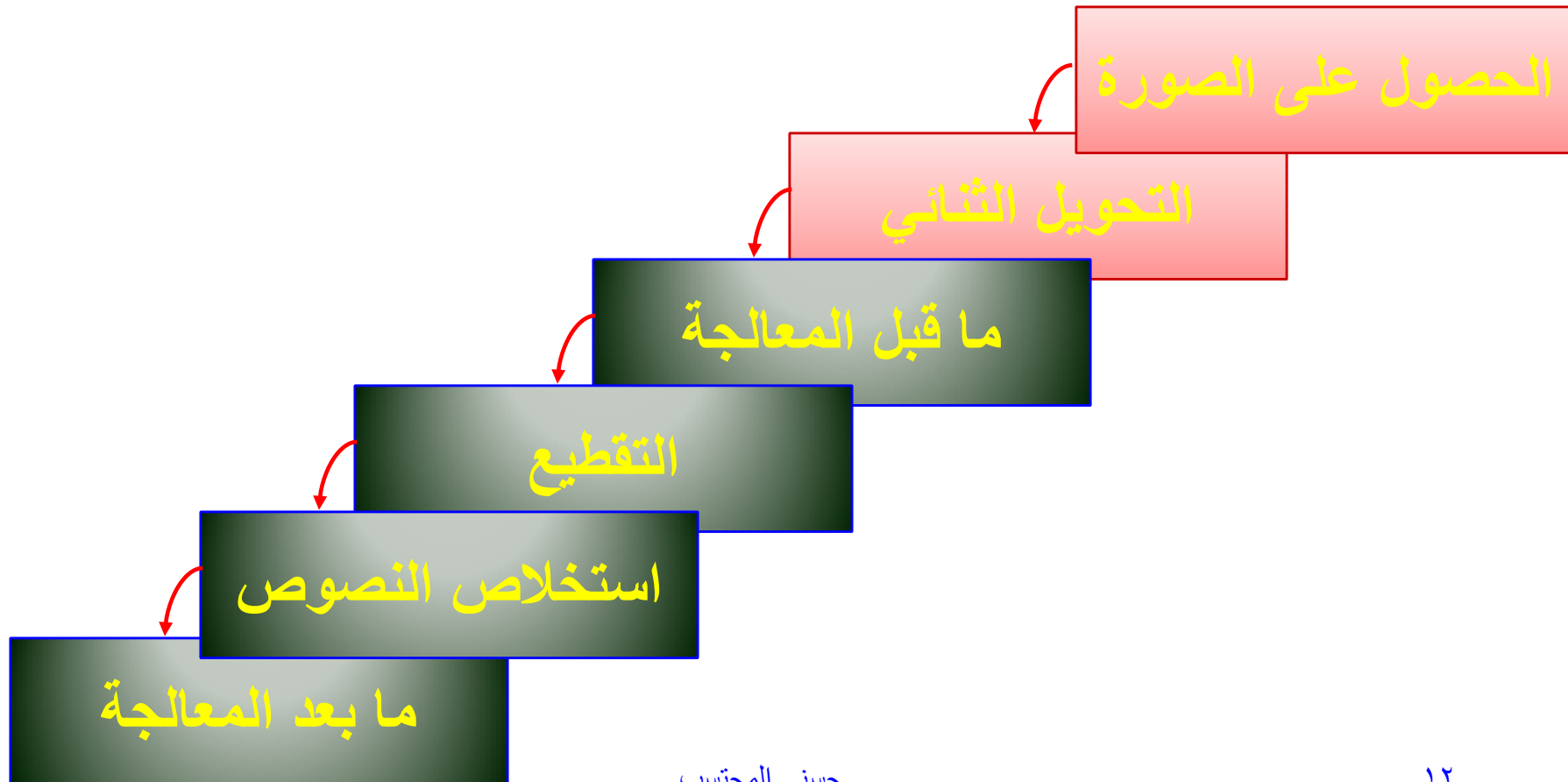
```
000 000 000 00 000 000 00 000 000 00 000 000 00 000 000 00 000 000
0000 000 000 01 000 000 01 110 011 00 000 000 00 000 001 00 000 000 11 000 00
000 000 111 11 100 000 01 110 000 00 000 000 00 000 001 10 000 000 11 100 000
000 000 111 11 110 000 01 110 000 11 100 000 00 000 001 10 000 000 11 100 000
000 000 111 11 100 000 01 110 000 11 110 000 00 000 001 10 000 000 01 100 000
000 000 010 00 000 000 01 111 100 11 111 100 00 000 001 10 000 000 11 011 000
000 000 000 00 000 000 11 111 110 11 111 110 00 000 001 10 000 001 11 111 100
000 000 000 00 000 000 11 111 100 11 111 111 10 000 001 11 000 001 11 111 100
000 000 000 00 000 000 11 101 100 01 111 111 11 000 001 11 000 001 11 011 000
000 000 000 01 100 000 00 000 000 00 111 011 11 000 001 11 000 000 10 000 000
000 000 000 01 100 000 00 000 000 00 000 001 11 100 001 11 000 000 00 000 000
000 000 000 11 100 000 00 000 000 00 000 000 01 110 001 11 000 000 00 000 000
000 000 001 11 110 000 00 000 000 00 000 000 00 111 011 11 000 000 00 000 000
000 000 011 11 110 000 00 111 000 00 000 000 00 111 011 11 100 000 00 100 000
000 000 111 01 110 000 00 111 000 00 000 000 00 011 111 11 100 000 01 110 000
000 000 110 01 110 000 00 111 000 00 000 000 00 001 110 01 100 000 01 110 000
000 001 111 11 110 000 00 111 000 00 000 000 00 001 110 01 100 000 01 110 000
000 001 111 11 110 000 00 111 000 00 000 000 00 001 110 01 110 000 00 110 000
000 001 111 11 111 000 00 111 000 00 000 000 00 001 100 01 110 000 01 110 000
000 000 111 00 111 111 11 110 000 00 000 000 00 011 100 00 111 001 11 110 000
000 000 000 00 011 111 11 110 000 00 000 000 00 111 000 00 111 111 11 100 000
000 000 000 00 001 111 11 100 000 00 000 000 11 110 000 00 011 111 11 000 000
000 000 000 00 000 011 00 000 000 00 000 001 11 100 000 00 001 111 10 000 000
000 000 000 00 000 000 00 000 000 00 000 111 11 000 000 00 000 000 00 000 000
000 000 000 00 001 110 00 000 000 00 001 100 00 000 000 00 000 000 00 000 000
```

الترقيم



ما قبل المعالجة

التعرف الضوئي على الكتابة العربية – طريقة العمل



ما قبل المعالجة

- تنقية الصورة من الشوائب
- ازالة الحواف السوداء المحيطة بالصورة والصور المرافقة للنص
- ازالة أي زوائد موجودة في النص كالأخطاء الناتجة من التصوير
- تصحيح الالتواءات
- تقسيم الأسطر
- تطبيع خطوط النص
- تقسيم الكلمات
- استخراج النقاط
- تنعيم هياكل الكلمات أو التنحيف

ما قبل المعالجة

- من العوامل المؤثرة في مستوى وضوح الصورة المأخوذة عن طريق الماسح الضوئي:
 - تاريخ الاصل أو قدم الورقة
 - جودة الورق
 - طريقة الطباعة
 - ليزر
 - نقطية
 - آلة كاتبة
 - سلامة الاصل من التظليل أو التخطيط
 - وضوح الخط
 - حجم الخط

ما قبل المعالجة - تعديل الحروف

- تقدير وتصحيح ميلان الحروف وتحديد الحروف بحجم معين
- تختلف الانظمة في طريقة تعاملها مع الحرف العربي

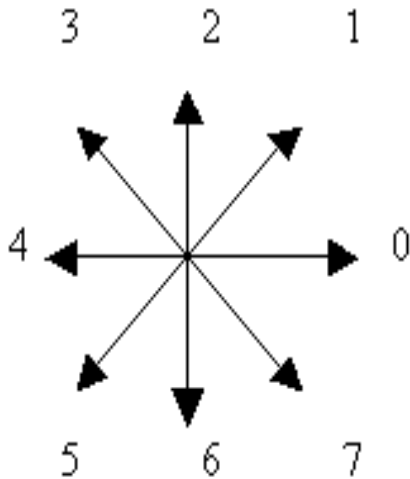
- مثلا

- معاملة ي وى كحرف واحد

- معاملة الحروف أ إ آ الى كحرف ا

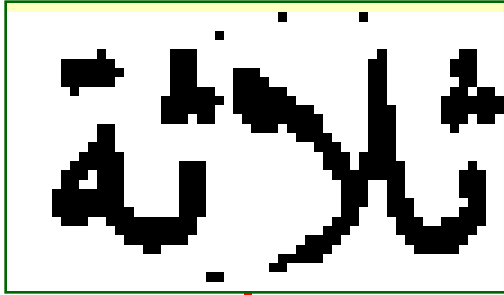
ما قبل المعالجة – التنعيم (الملاسة)

- يمكن تحديد حالة نقطة معينة (pixel) من خلال الاستفادة من النقاط الثمان المحيطة بها
- تشمل عملية الملاسة إزالة التشويش والأخطاء بالإضافة الى عملية التنحيف



ما قبل المعالجة – إزالة التشويش

- تتضمن عملية تنقية الصورة من الشوائب إزالة جميع الحواف السوداء المحيطة بالصورة والصور المرافقة للنص كما تتضمن إزالة أي زوائد موجودة في النص كالأخطاء الناتجة من التصوير

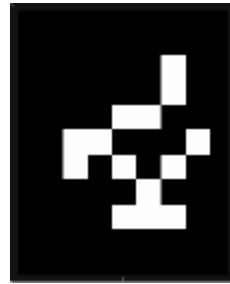


ما قبل المعالجة – إزالة التشويش

- تعتبر عملية إزالة التشويش في اللغة العربية غير سهلة بسبب كثرة النقاط ووجود حروف منفصلة
- قبل الحكم على نقطة بأنها من الشوائب
 - وجود النقطة السوداء
 - النقطة غير معزولة
 - النقطة محاطة من الجوانب الأربعة

ما قبل المعالجة – إزالة التشويش

• تنقية الاجزاء المتصلة قد يزيل الاجزاء المعزولة



• تنقية المحيط: ويقوم بعملية ازالة للتشويش المتصل والثقيل.



ما قبل المعالجة - التنحيف

- تتم عملية التنحيف وذلك بحذف النقاط العرضية الزائدة الواصلة بين نقطتين، فمثلا قد تكون الألف عبارة عن مستطيل ذو عرض معين فيحول إلى مستقيم عرضه نقطة واحدة
- حذف النقاط الزائدة وحفظ أقل عدد من النقاط الضرورية للتعرف على الحرف يساعد في تسريع الخطوات اللاحقة وتوفير الذاكرة
- الهدف الاساسي من التنحيف إيجاد هيكل الحروف.
- مثال:

المسافة في

المسافة في

ما قبل المعالجة - التنحيف

- وجود النقاط المفصولة من القضايا الحساسة في عملية التنحيف ولذلك لا بد من إزالتها قبل إجراء عملية التنحيف
- ينتج عن عملية التنحيف ايجاد هيكل الحرف
- يعاب على عملية التنحيف:
 - بطء عملية المعالجة
 - ظهور هياكل جديدة واختفاء أخرى

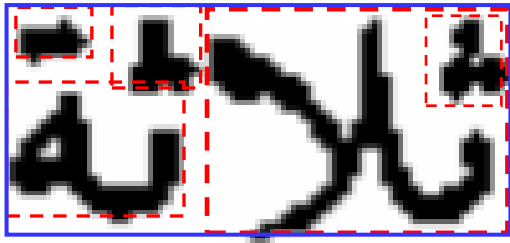


ما قبل المعالجة - التثخيف

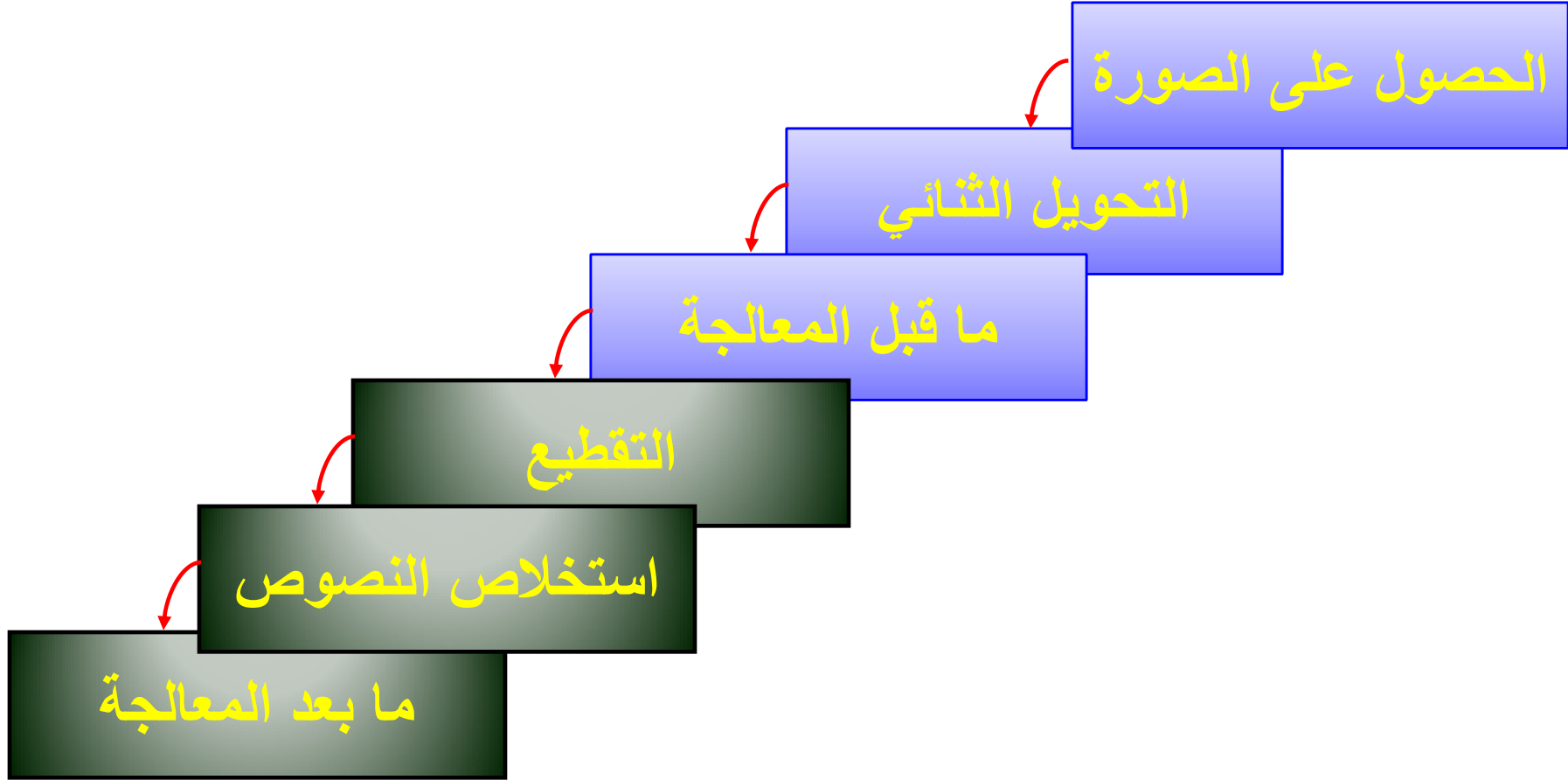
- بعض الخوارزميات المستخدمة في عملية التثخيف
 - التثخيف المتسلسل
 - التثخيف المتوازي
 - التثخيف غير المتكرر

ما قبل المعالجة – الأجزاء المتصلة

- قبل عملية التقطيع يجب إجراء عملية التهيئة والتي تشترط وجود الكلمات في سطور
- الكلمات العربية تتكون من حروف متصلة وتعتمد أشكال الحروف على مكانها في الكلمة مما يصعب عملية التقسيم
- تقسيم الكلمات إلى أجزاء وتقسيم الكلمة نفسها إلى أجزاء أصغر بحيث تكون الحروف المتصلة في جزء والمعزولة في جزء آخر



التعرف الضوئي على الكتابة العربية – طريقة العمل



التقطيع

- فصل الكلمات والحروف لتسهيل عملية التعرف
- صعوبة مقارنة بالانجليزية
- من أهم الخطوات

غزت عقول



غزت عقول

التقطيع - الأخطاء

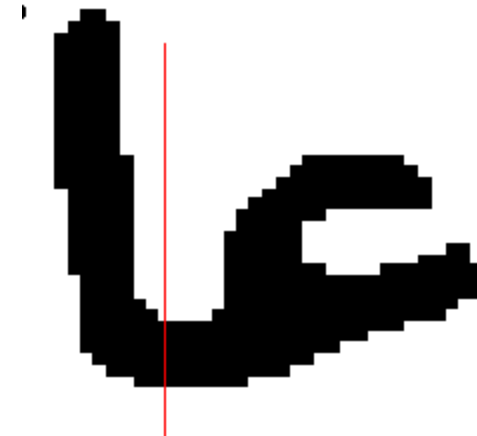
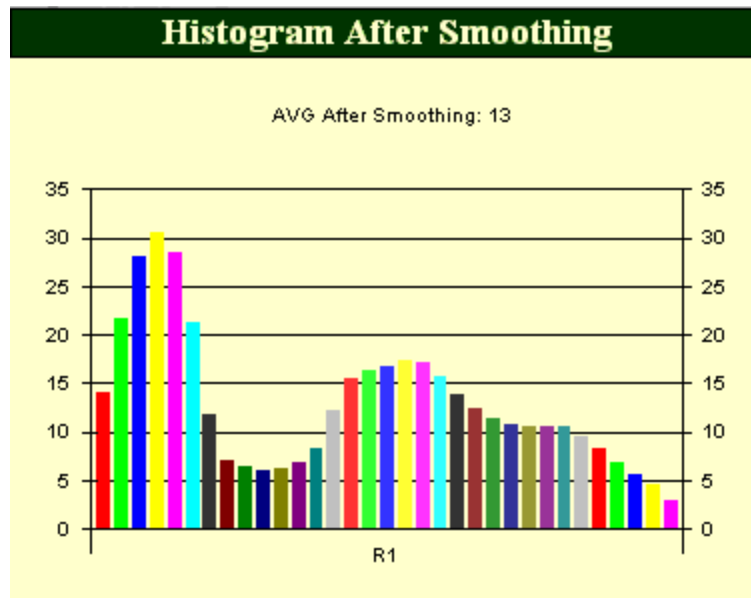
- التقطيع أكثر من اللازم
- التقطيع أقل من اللازم
- خطأ في أماكن التقطيع

سطح

التقطيع – طرق التقطيع

- المدرج الإحصائي للإسقاط العمودي
- استخدام الحد الأعلى المحلي كنقطة قطع
- إذا كان طول العمود أقل من المتوسط / ٣ يعتبر نقطة قطع

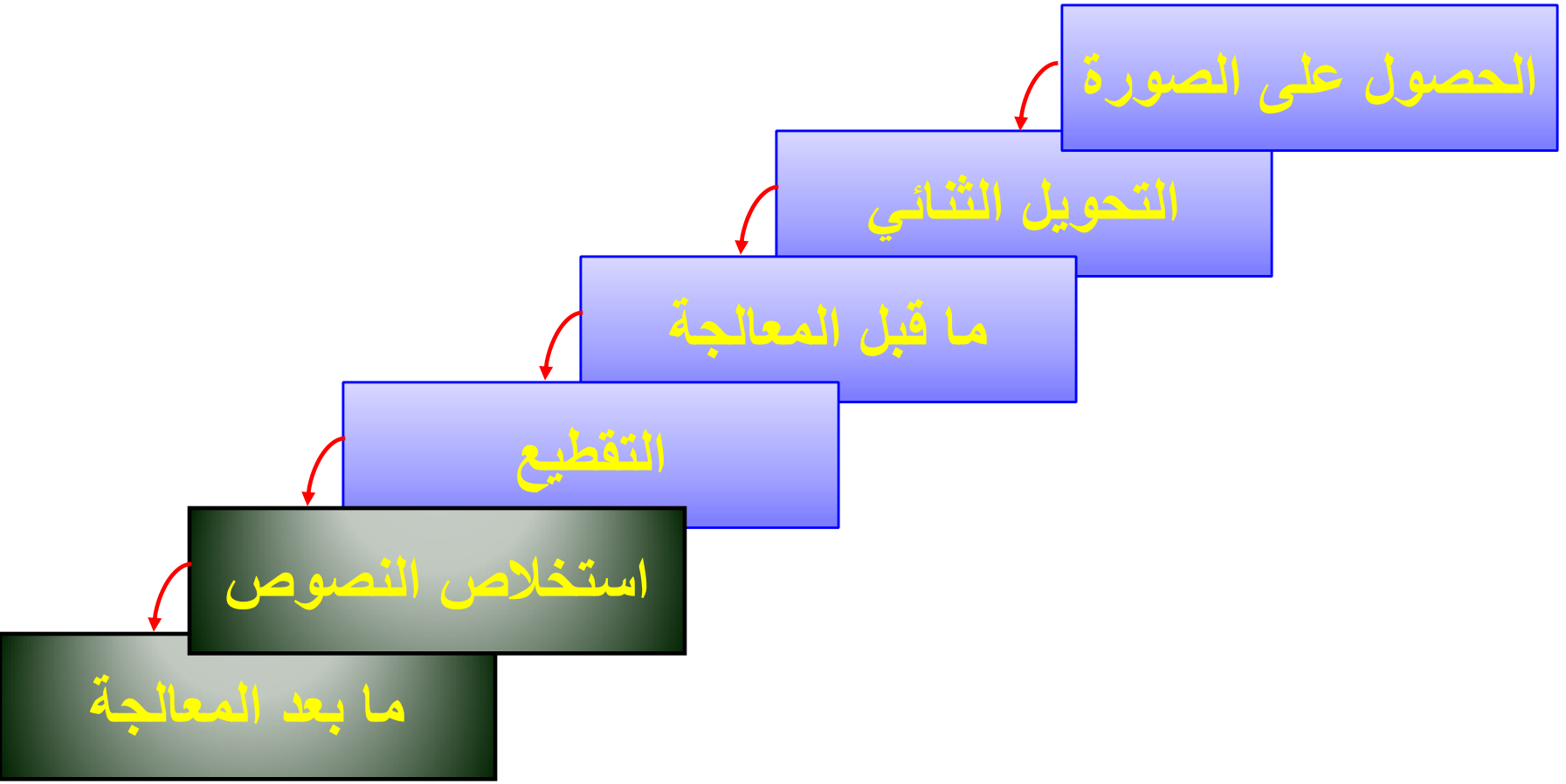
التقطيع – طرق التقطيع



التقطيع – طرق التقطيع

س

التعرف الضوئي على الكتابة العربية – طريقة العمل



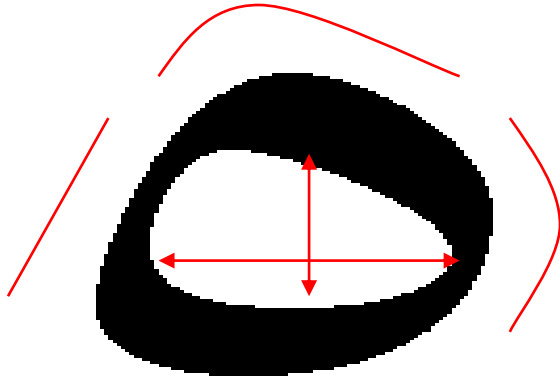
استخلاص الخصائص

- بعد التقطيع تستخلص الخصائص المميزة لكل شكل
- تجمع الأشكال إلى مجموعات وتعطى كل مجموعة رمز فمثلا

نوع الشكل	الرمز
حرف	١
جزء من حرف	٢
مجموعة من حروف	٣
نقطة عليا	٤

استخلاص الخصائص

• الخصائص التركيبية



استخلاص الخصائص

• الخصائص الإحصائية

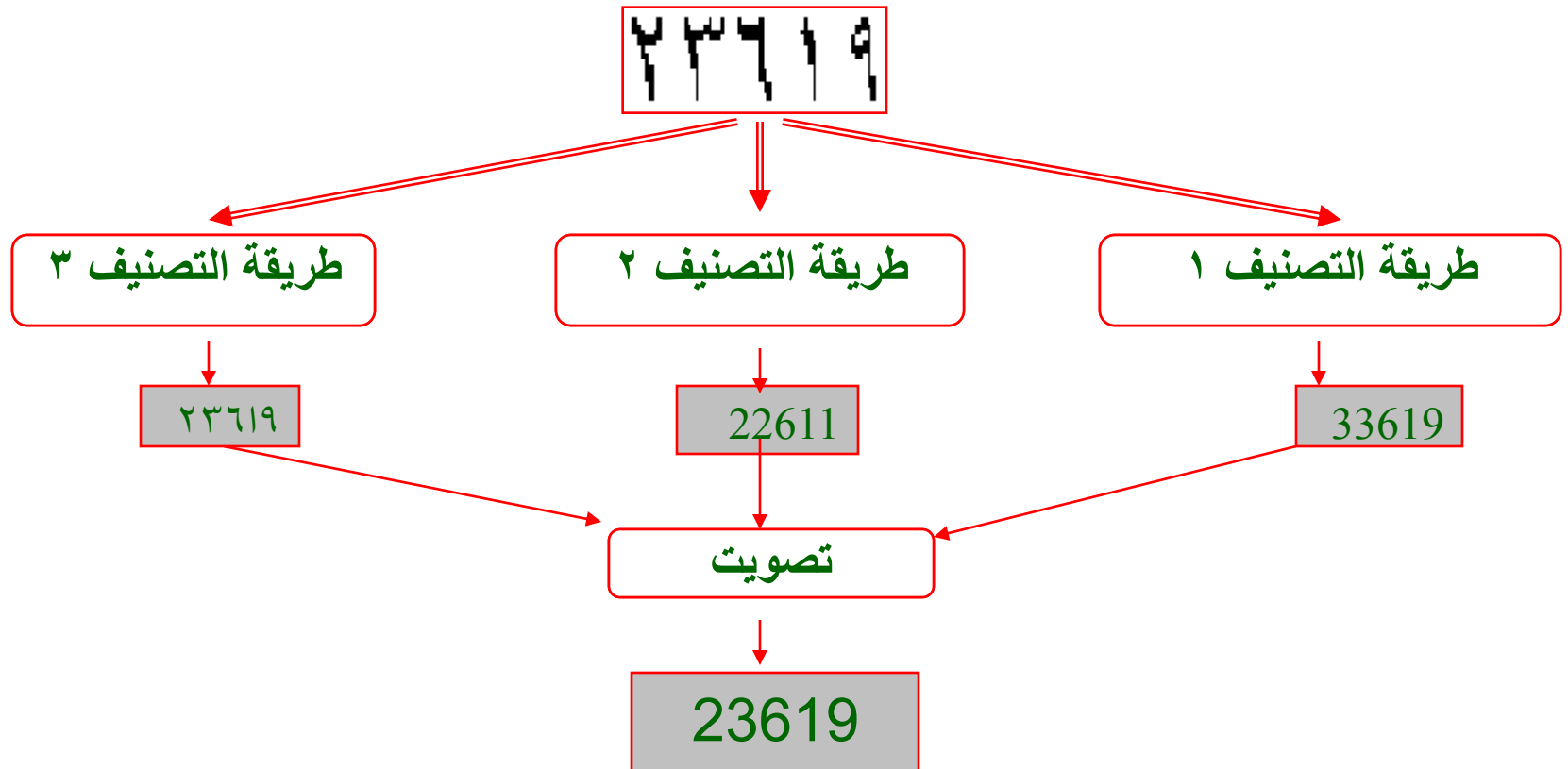
- الطول
- العرض
- الكثافة في الاتجاه الرأسي
- الكثافة في الاتجاه الأفقي
- عدد التقاطعات في الاتجاه الرأسي
- عدد التقاطعات في الاتجاه الأفقي
- نسبة النقاط البيضاء إلى السوداء



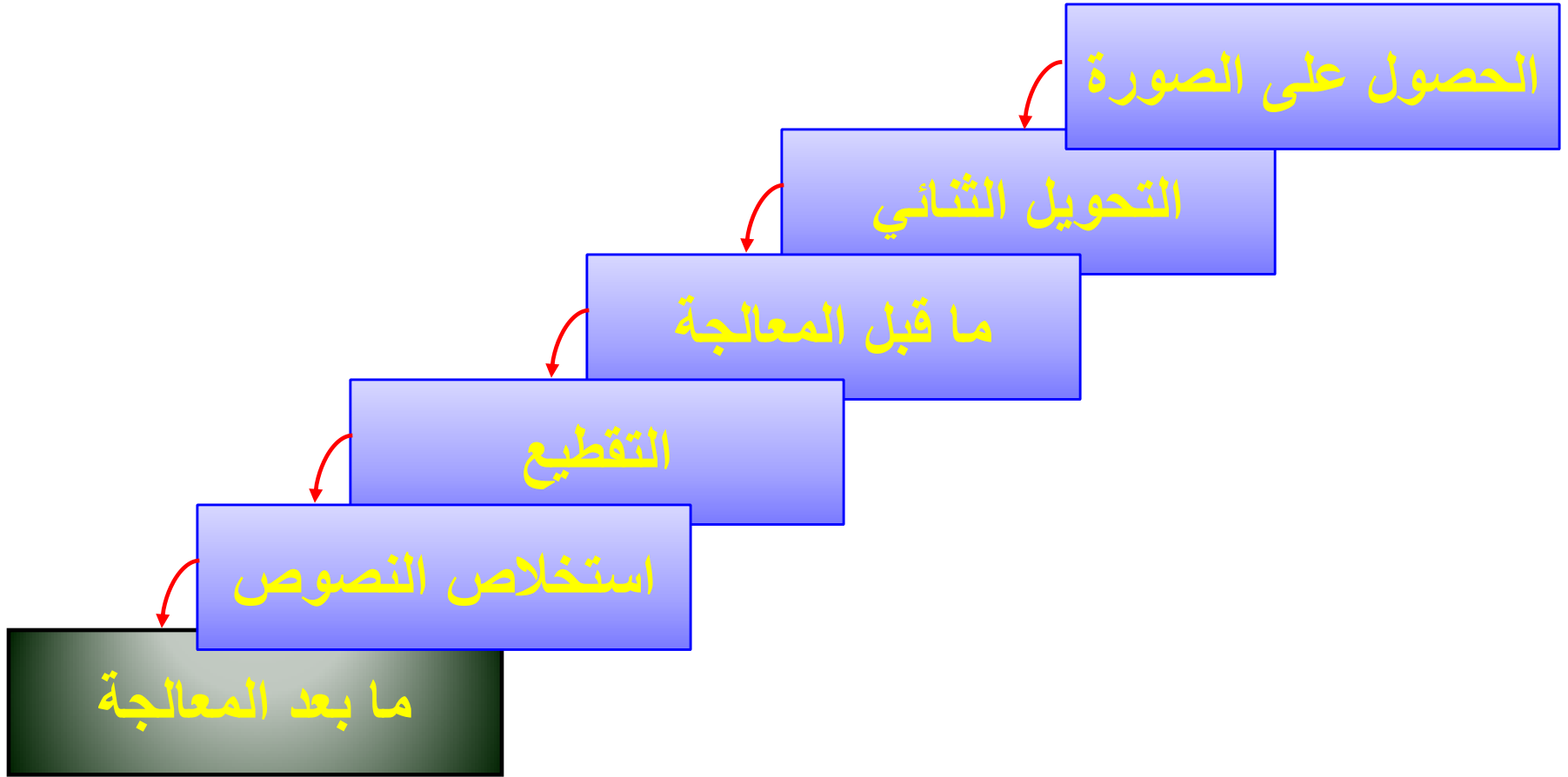
التصنيف

- التصنيف هو عملية تحديد الحرف
- المقارنة
- التصويت

التصنيف



التعرف الضوئي على الكتابة العربية – طريقة العمل



ما بعد المعالجة

- التدقيق الإملائي وتصحيح الأخطاء الناتجة عن عمليتي ما قبل المعالجة والمعالجة
- مثلاً التأكد من صحة الكلمة بمطابقتها للكلمات الموجودة في معجم مصاحب
- عند عدم وجود الكلمة قد يسأل المستخدم للتصحيح أو إضافتها للمعجم

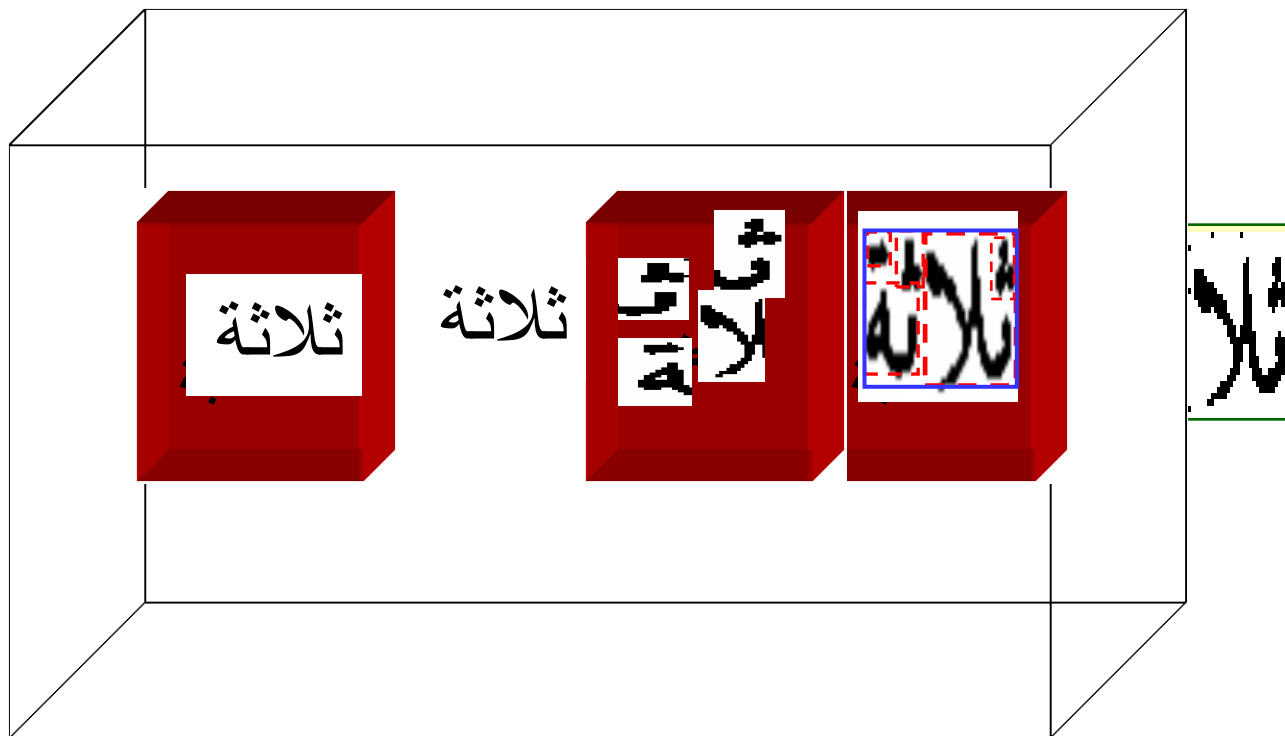
مقارنة

- طريقة الصندوق الاسود



مقارنة

- طريقة الصندوق الابيض



مقارنة

- القارئ الآلي
- OmniPage

عناصر معايير المقارنة

- عناصر المعايرة
 - نص: عدد الرموز الموجودة في النص الأصلي
 - ناتج: عدد الرموز الموجودة في النص الناتج عن عملية التعرف على الكتابة العربية
 - صح: عدد الرموز المتعرف عليها بشكل صحيح
 - مس: عدد الرموز المستبدلة
 - مح: عدد الرموز المحذوفة
 - مض: عدد الرموز المضافة

معايير المقارنة

- **الضبط:** عدد الرموز المتعرف عليها بشكل صحيح على عدد الرموز الموجودة في النص الأصلي

$$\text{الضبط} = \text{صح} / \text{نص}$$

- **الدقة:** عدد الرموز المتعرف عليها بشكل صحيح على عدد الرموز الموجودة في النص الناتج عن عملية التعرف على الكتابة العربية

$$\text{الدقة} = \text{صح} / \text{ناتج}$$

- **المضاف:** عدد الرموز المضافة على عدد الرموز الموجودة في النص الأصلي

$$\text{المضاف} = \text{مض} / \text{نص}$$

معايير المقارنة

• **المحذوف:** عدد الرموز المحذوفة على عدد الرموز الموجودة في النص الأصلي

المحذوف = مح / نص

• **المستبدل:** عدد الرموز المستبدلة على عدد الرموز الموجودة في النص الأصلي

المستبدل = مس / نص

المقارنة

انني اكتب لكم هذه الانطباعات من بيت هادىء من بيوت مدينة هادئة أمنة كبقية مدن الوطن.. أكثر ما يسبب الازعاج المتكرر هو اجراس الباب.. والهاتف ذو الخطوط المتداخلة.. والسيارة غير المتكافئة مع مهماتها العديدة.. وليس اصوات قنابل وأصداء جرائم

OmniPage

انني اكتب لكم هذا الانطباعات من بيت هادىء من بيوت مدينة هادئة أمنة كبقية مدن الوطن . أكثرها يسبب الازعاج المتكرر هو اجراس الباب . والهاتف ذو الخطوط المتداخلة . والسيارة غير المتكافئة مع مهماتها العديدة . وليس اصوات قنابل وأصداء جرائم

94.06 %

القارئ الآلي

لم نثني /كتب لكم هذه لم لانطباعات من بليت مادىء من بليبرت مدلية هادئة لم منة كبقية مدن لم لوطن . . لم كثرما لشيبب /الازعاج لم لتكرر هو /جرلم س لم لباب . . لم لهاتف ذو لم لخطوط /التدخلة . . و/السيارة غير لم لمتكافئة مع مهاتها ،لعدب ة . .بميم لم صؤت قنابل ولم صدمء جزئم

83.67 %

المقارنة

السياسة أن الشعب اللبناني ذاته ، في الاستفتاء الذي جرى في شهر حزيران الماضي ، عبّر عن تبنيه لها بالذات وعن إيمانه بصوابها ، فأصبح من واجب الحكومة المنبثقة من مجلسكم الكريم - وهو ثمرة هذا الاستفتاء الشعبي - أن تبقى أمينة لتلك السياسة وأن تستمر في تنفيذها .

OmniPage

اليامة أن الشعب اللبناني ذاته ، في الامتفتا ،لذي جرى في شهر حزيران الماني ،
عبر عن تبنيه لها بالذات وي
إيأنه بعوام 1 ، فأصبح ء . واجب المكومة المنبثقة من مجلاا الكرم _ وهو ثرة هذا
الاستفتاء الشعبي _ أن تبق
أمينة لتلك اليات وأن تتمو في تنفيذها .

89.14 %

القارئ الآلي

السماسة أر السصط اللسايبى ذاته ، بي الاسصصاء الذي جرى ي شهر حريراد المامي ، ككمر س تبسبه لها بالذات وكل
إيمأنه لصواكا ، لاصبح ص ط واجط ألحكومة المشمة س محلسم الكريم -وهو ثمرة هذا الاسصصعاء الاعى - أن تبقى
أمسة لمللت السماسة وأر نسمر ي تمدها .

76.3 %

المقارنة

الجزائر، لم يكن الجزائريون يعتبرون
انفسهم فرنسيين، ولهذا بدأت الثورة
الجزائرية ضد فرنسا وقادها سياسيون

OmniPage

! ل نل
نفسم في نسيين ، و بد رة
ية ضد في و □ ن

41.09 %

القارئ الآلي

الجزائر، لم يكن الجزائريون يعتبرون
انفسهم فرنسيين ، ولهذا بدأت الثورة
الجزائرية ضد فرنسا وقادها سياسيون

98.13 %

المقارنة

المرحلة ينبغي ان تدرسها الدول العربية بدقة تامة، فهي من النوع الذكي الذي تحبكه الصهيونية بمهارة فائقة والذي يعكس شعور اسرائيل بالخطر من تنامي مد التواصل السياسي والدبلوماسي بين اميركا والدول العربية في

OmniPage

المرحلة ينبغي ان تدرسها الدول العربية بدقة تامة ، فهي من النوع الذكي الذي تحبكه الصهيونية بمهارة فائقة والذي يعكس شعور اسرائيل بالخطر من تنامي مد التواصل السياسي والدبلوماسي بين اميركا والدول العربية في

97.7 %

القارئ الآلي

المرحلة ينبغي ان تدرسها الدول العربية بدقة تامة، فهي من النوع الذكي الذي تحبكه الصهيونية بمهارة فائقة والذي يعكس شعور اس ائيل بالخطر من تنامي مد التواصل السياسي والدبلوماسي بين اميركا والدول العربية في

98.08 %

المقارنة

• معدل دقة الصفحة

• القارئ الآلي ٩٠.٣٣٣%

• OmniPage ٨٦.٨٩%

خاتمة

- مقدمة
- طريقة العمل
- التحويل الثنائي
- ما قبل المعالجة
- التقطيع
- استخلاص الخصائص والتصنيف
- ما بعد المعالجة
- مقارنة

وشكرا

سبحانك اللهم وبحمدك أشهد أن لا إله
إلا أنت أستغفرك وأتوب إليك