Page Segmentation in Urdu Nastalique OCR

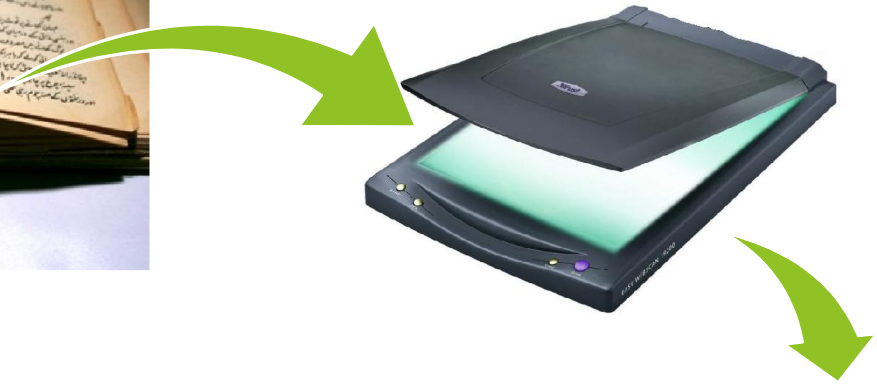
Qasim Ali
qasim.ali@kics.edu.pk

29-September, 2014

مرکز تحقیقات لسانیات

Center for Language Engineering
Al- Khwarizmi Institute of Computer Sciences
University of Engineering and Technology, Lahore, Pakistan

Optical Character Recognition

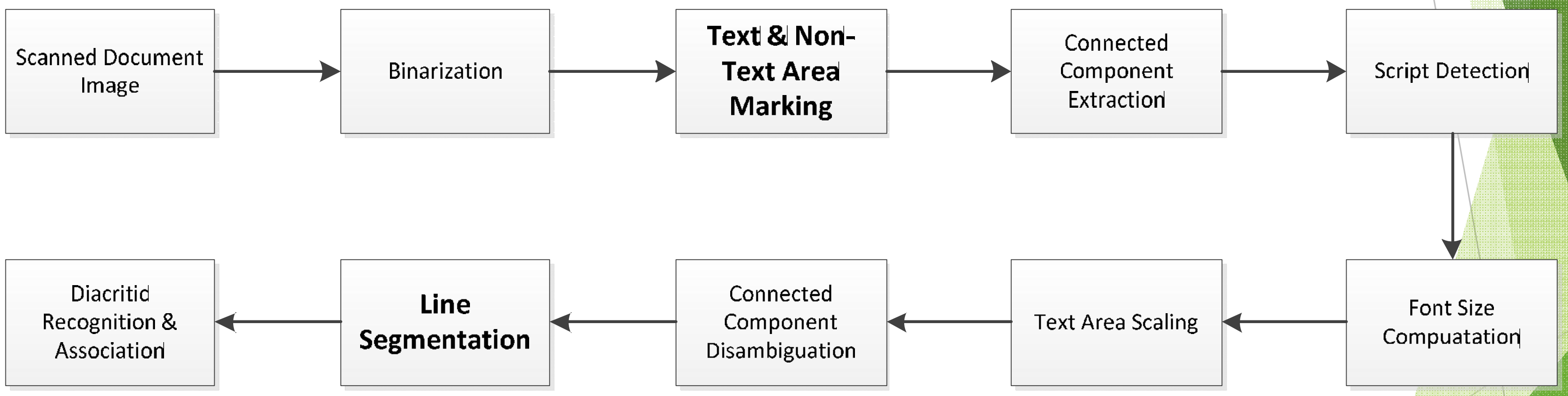


106

آئے ہوئے تمام لوگوں کا شکریہ ادا کرتا ہے اور مرحوم یا مرحومہ کی طرف سے معافی کا
طلبگار ہوتا ہے کہ حیات میں ان سے کوئی غلطی ہوئی ہو، وہ معاف کر دیں اور خدا سے
دعا کرو۔ پھر مجلس برخاست ہوتی ہے۔ سارے لوگ اپنے اپنے گھروں کی طرف چلے
جاتے ہیں ابل ہمسائے اور برادری کے لوگ اس موت کے سر چلے جاتے ہیں، جہاں
پر ہمسائے ان کے لیے چائے اور اس کے ساتھ روٹیاں لے آتے ہیں، وہ پیش کی جاتی
ہیں۔ وہاں پھر فاتحہ انی ہوتی ہے۔ اس دوران یہ سارے رسوم خلیفہ کے ذریعے ادا
ہوتے ہیں، جہاں پر خلیفہ ہر وقت ان کے درمیان موجود رہتا ہے اور جو لوگ فاتحہ خوانی
کے لیے آتے ہیں املنا کے لیے فاتحہ خوانی کراتا ہے۔

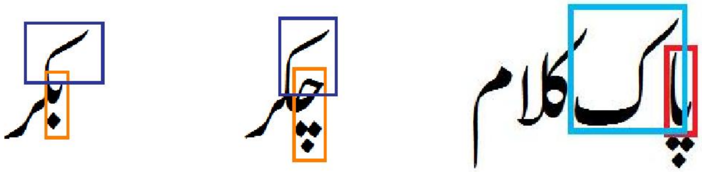
Phases In OCR

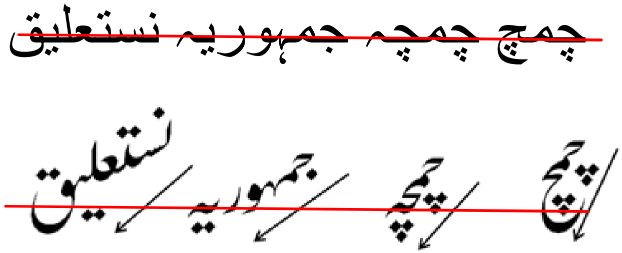
1. Pre-processing

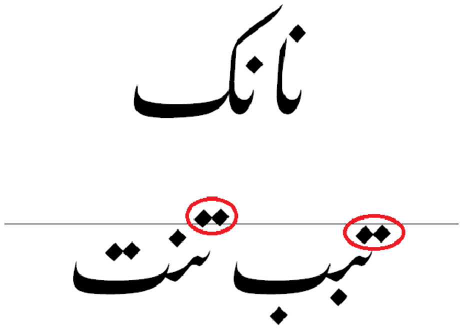


2. Classification and recognition
3. Post-processing

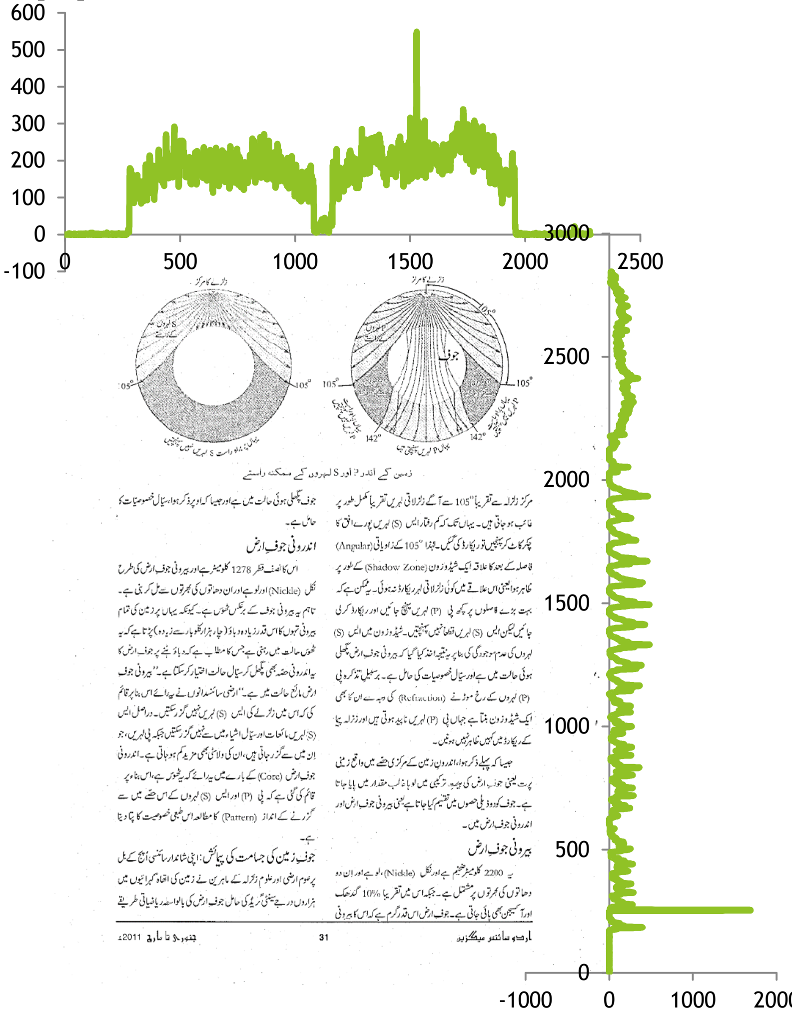
Nastalique Writing Style Complexities







Top Down Approach



زمین کے اندر P اور S لہروں کے ممکنہ راستے

مرکز زلزلہ سے تقریباً 105° سے آگے زلزلاتی لہریں تقریباً مکمل طور پر غائب ہو جاتی ہیں۔ یہاں تک کہ کم رفتار ایس (S) لہریں پورے افق کا چکر کاٹ کر پہنچیں تو ریکارڈ کی گئیں۔ لہٰذا 105° کے زاویاتی (Angular) فاصلہ کے بعد کا علاقہ ایک شیڈو زون (Shadow Zone) کے طور پر ظاہر ہوا یعنی اس علاقے میں کوئی زلزلاتی لہر ریکارڈ نہ ہوئی۔ یہ ممکن ہے کہ بہت بڑے فاصلوں پر کچھ پی (P) لہریں پہنچ جائیں اور ریکارڈ کر لی جائیں لیکن ایس (S) لہریں قطعاً نہیں پہنچتیں۔ شیڈو زون میں ایس (S) لہروں کی عدم موجودگی کی بنا پر یہ نتیجہ اخذ کیا گیا کہ بیرونی جوف ارض پگھلی ہوئی حالت میں ہے اور سیال خصوصیات کی حامل ہے۔ بر سبیلِ تذکرہ پی (P) لہروں کے رخ موڑنے (Refraction) کی وجہ سے ان کا بھی ایک شیڈو زون بنتا ہے جہاں پی (P) لہریں ناپید ہوتی ہیں اور زلزلہ پیما کے ریکارڈ میں کہیں ظاہر نہیں ہوتیں۔

جیسا کہ پہلے ذکر ہوا، اندرونِ زمین کے مرکزی حصے میں واقع زمینی پرت یعنی جوفِ ارض کی ہیئتِ ترکیبی میں لوہا غالب مقدار میں پایا جاتا ہے۔ جوف کو دو ذیلی حصوں میں تقسیم کیا جاتا ہے یعنی بیرونی جوفِ ارض اور اندرونی جوفِ ارض میں۔

بیرونی جوفِ ارض

یہ 2200 کلومیٹر ضخیم ہے اور نکل (Nickle)، لوہے اور ان دو دھاتوں کی بھرتوں پر مشتمل ہے۔ جبکہ اس میں تقریباً 10% گندھک اور آکسیجن بھی پائی جاتی ہے۔ جوف ارض اس قدر گرم ہے کہ اس کا بیرونی جوف پگھلی ہوئی حالت میں ہے اور جیسا کہ اوپر ذکر ہوا، سیال خصوصیات کا حامل ہے۔

اندرونی جوفِ ارض

اس کا نصف قطر 1278 کلومیٹر ہے اور بیرونی جوفِ ارض کی طرح نکل (Nickle) اور لوہے اور ان دھاتوں کی بھرتوں سے مل کر بنی ہے۔ تاہم یہ بیرونی جوف کے برعکس ٹھوس ہے۔ کیونکہ یہاں پر زمین کی تمام بیرونی تہوں کا اس قدر زیادہ دباؤ (چار ہزار کلوبار سے زیادہ) پڑتا ہے کہ یہ ٹھوس حالت میں رہتی ہے جس کا مطلب ہے کہ دباؤ ہٹنے پر جوف ارض کا یہ اندرونی حصہ بھی پگھل کر سیال حالت اختیار کر سکتا ہے۔ "بیرونی جوف ارض مائع حالت میں ہے۔" ارضی سائنسدانوں نے یہ رائے اس بنا پر قائم کی کہ اس میں زلزلے کی ایس (S) لہریں نہیں گزر سکتیں۔ دراصل ایس (S) لہریں مائعات اور سیال اشیاء میں سے نہیں گزر سکتیں جبکہ پی لہریں، جو ان میں سے گزر جاتی ہیں، ان کی ولاسٹی بھی مزید کم ہو جاتی ہے۔ اندرونی جوفِ ارض (Core) کے بارے میں یہ رائے کہ یہ ٹھوس ہے، اس بناء پر قائم کی گئی ہے کہ پی (P) اور ایس (S) لہروں کے اس حصے میں سے گزرنے کے انداز (Pattern) کا مطالعہ اس طبعی خصوصیت کا پتا دیتا ہے۔

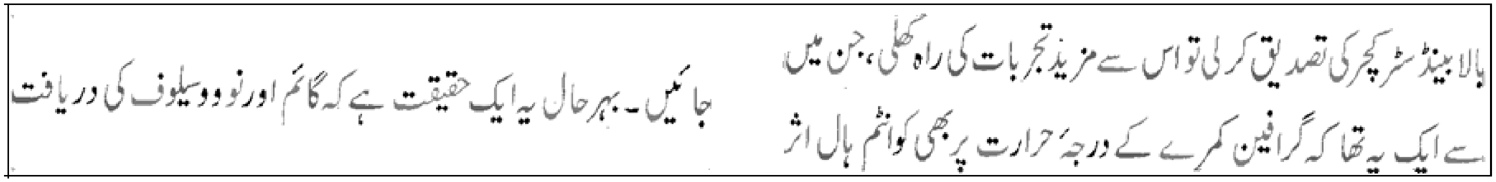
جوفِ زمین کی جسامت کی پیمائش: اپنی شاندار سائنسی اپج کے بل پر علوم ارضی اور علوم زلزلہ کے ماہرین نے زمین کی اتھاہ گہرائیوں میں ہزاروں درجے سینٹی گریڈ کی حامل جوف ارض کی بالواسطہ ریاضیاتی طریقے

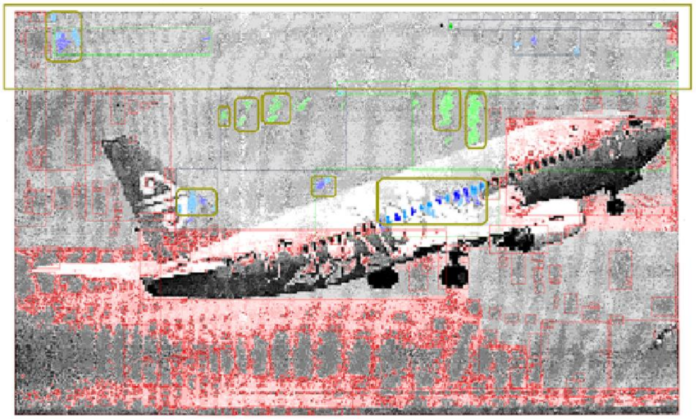
Bottom Up Approach

9

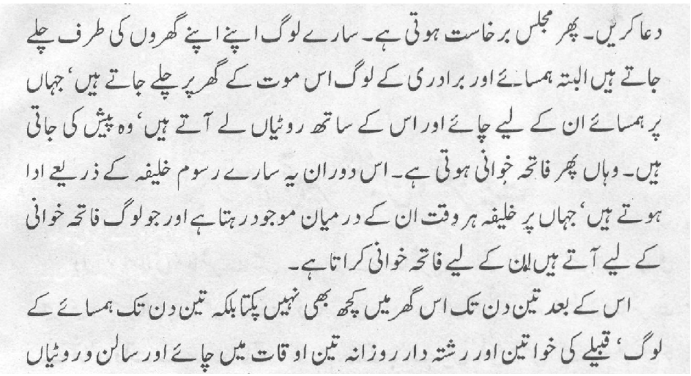
"(اے رسول) انہیں کہہ دیجیے (چشمِ بصیرت لے کر) چل پھر کر امم گزشتہ کے ادبار و
زوال کے کھنڈرات دیکھو اور مجرموں کے انجام سے (درسِ عبرت لو)
(سورہ النمل 27: ایت 69)
ماضی میں جھانکنے یا کھوج لگانے کے لیے علم آثار قدیمہ نے جو کہ اب ایک اہم سائنس
ہے، نہایت اہم کام کیا ہے۔ علم آثار قدیمہ کو ایک جذباتی سائنس کہا جاتا ہے کیونکہ اس میں
بے آرزو ہیجان اور بہت زیادہ تکلیف دہ سائنسی طریقے شامل ہیں۔ پوری دنیا میں یہ سائنس
اٹھارویں صدی میں پوم پی آئی (Pompeii) کی کھدائی کے ساتھ معرض وجود میں آئی۔
پوم پی آئی شہر اٹلی کے جنوب مغرب میں واقع تھا۔ جو 79ء میں وسوویئس
(Visuvius) پہاڑ کے پھٹنے سے آتش فشاں کے ملبے کے نیچے دفن ہو گیا تھا۔ 90 فی صد شہر
کی آبادی وقت میں باہر آ گئی تھی لیکن 2000 لوگوں نے شہر چھوڑنے سے انکار کر دیا تھا۔ ہلکی
مسام دار چٹان جسے پومک پتھر کہتے ہیں، اس کے چاولوں کے دانوں کے برابر اور دوسرے
مشت بھر جسامت کے 15 سنٹی میٹر (6 انچ) فی گھنٹہ کی رفتار سے بارش کی طرح گلیوں میں
گرے اور پھر اسی پہاڑ سے ایک بادل خاک گیس اور چٹان پر مشتمل نیچے آیا اور اس دم گھٹنے
والے میٹیریل کے طوفان سے باقی ماندہ پوم پی آئی کے لوگ یا باشندوں میں سے کوئی نہ بچا۔
19 گھنٹوں کے اندر یہ شہر 25 فٹ (8 میٹر) موٹی طے کے ملبے کے نیچے دفن ہو چکا تھا جو
سیمنٹ کی طرح سخت ہو گئی۔ اس دفن شدہ شہر سے 1700 سال تک کسی نے کوئی تعلق نہ رکھا،
پھر 1863 میں ماہرین آثار قدیمہ نے پوم پی آئی کی کھدائی شروع کی۔
علمِ آثارِ قدیمہ کی ابتدائی نشوونما یا فروغ عہد جدید کی طلوعِ سحر کے ساتھ فنونِ لطیفہ
کے ماہرین کی رومانوی تحریک منطبق ہو گئی جو پہلے ہی جذبے اور تصور کی گہری سوچ میں غرق
تھے، ان کو بنجر و بیابان تصویر کی مانند اور خوشنما مقامات کی طرف لایا گیا جن میں کھنڈرات بھی
شامل تھے اس سے پہلے تو کوئی بھی اس کا استعمال نہ کرتا تھا بلکہ تاریخ کے اس ملبے میں کوئی
دلچسپی نہ لیتا تھا۔ علمِ آثار قدیمہ کو جس کا تعلق قدیم زمانے کے مطالعے سے ہے ابتدائی زندگی
میں ـــ کہیں سے جو بھی شہادت پائی جا سکی بھی حاصل کی۔ درحقیقت سائنس اور رومانس کی
دوھری مہک حاصل کی، جونہی لوگوں نے اپنے ماضی کے بارے میں کھوج لگانی شروع کی اور
مستقبل کے بارے میں سراغ لگانے کی امید میں اپنے ماضی کے بارے میں سوال کیا، تو اُن

Challenges In Two Column Segmentation



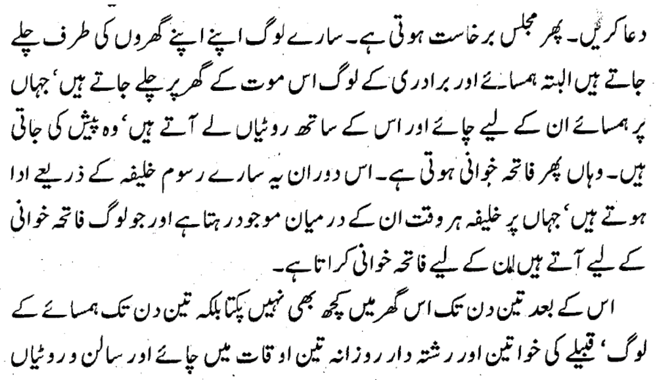


Layout Extraction



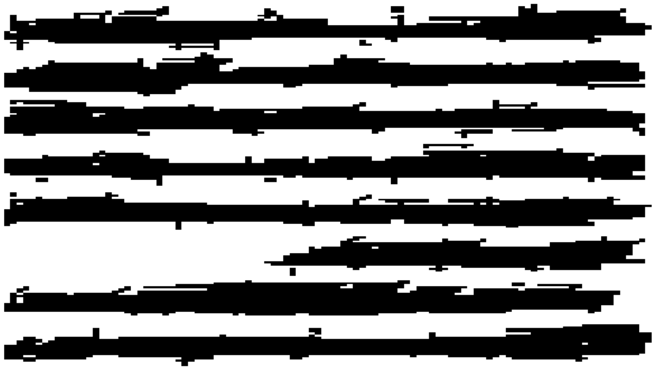
دعا کریں۔ پھر مجلس برخاست ہوتی ہے۔ سارے لوگ اپنے اپنے گھروں کی طرف چلے
جاتے ہیں البتہ ہمسائے اور برادری کے لوگ اس موت کے گھر پر چلے جاتے ہیں' جہاں
پر ہمسائے ان کے لیے چائے اور اس کے ساتھ روٹیاں لے آتے ہیں' وہ پیش کی جاتی
ہیں۔ وہاں پھر فاتحہ خوانی ہوتی ہے۔ اس دوران یہ سارے رسوم خلیفہ کے ذریعے ادا
ہوتے ہیں' جہاں پر خلیفہ ہر وقت ان کے درمیان موجود رہتا ہے اور جو لوگ فاتحہ خوانی
کے لیے آتے ہیں ان کے لیے فاتحہ خوانی کراتا ہے۔
اس کے بعد تین دن تک اس گھر میں کچھ بھی نہیں پکتا بلکہ تین دن تک ہمسائے کے
لوگ' قبیلے کی خواتین اور رشتہ دار روزانہ تین اوقات میں چائے اور سالن و روٹیاں

Scanned image of font size 14

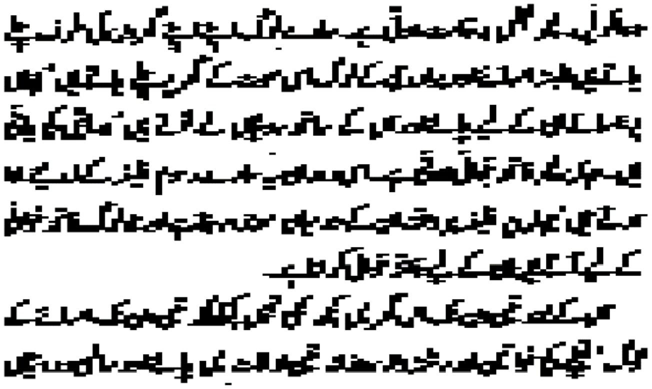


دعا کریں۔ پھر مجلس برخاست ہوتی ہے۔ سارے لوگ اپنے اپنے گھروں کی طرف چلے
جاتے ہیں البتہ ہمسائے اور برادری کے لوگ اس موت کے گھر پر چلے جاتے ہیں' جہاں
پر ہمسائے ان کے لیے چائے اور اس کے ساتھ روٹیاں لے آتے ہیں' وہ پیش کی جاتی
ہیں۔ وہاں پھر فاتحہ خوانی ہوتی ہے۔ اس دوران یہ سارے رسوم خلیفہ کے ذریعے ادا
ہوتے ہیں' جہاں پر خلیفہ ہر وقت ان کے درمیان موجود رہتا ہے اور جو لوگ فاتحہ خوانی
کے لیے آتے ہیں ان کے لیے فاتحہ خوانی کراتا ہے۔
اس کے بعد تین دن تک اس گھر میں کچھ بھی نہیں پکتا بلکہ تین دن تک ہمسائے کے
لوگ' قبیلے کی خواتین اور رشتہ دار روزانہ تین اوقات میں چائے اور سالن و روٹیاں

Binarized version

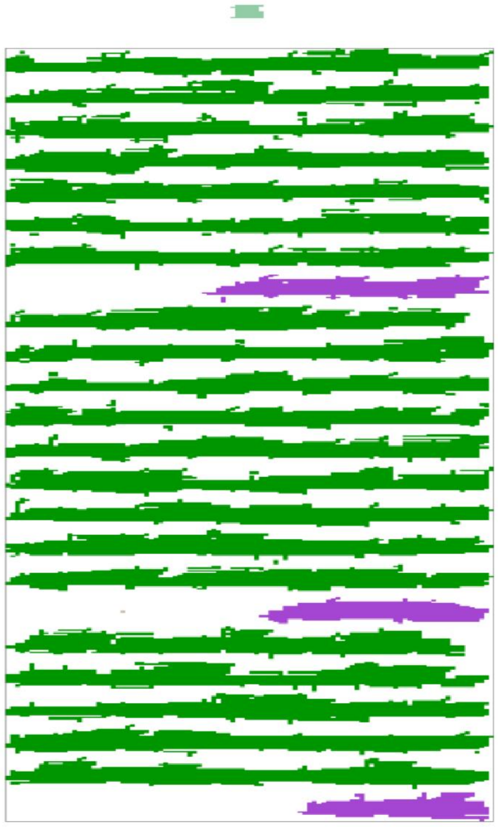


Resultant image after smearing

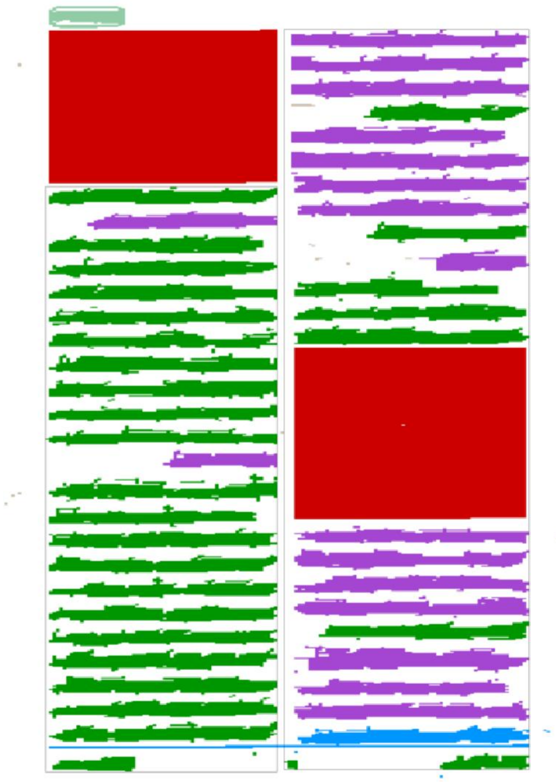


725x1306 pixels down sampled to 182x109

Layout Extraction

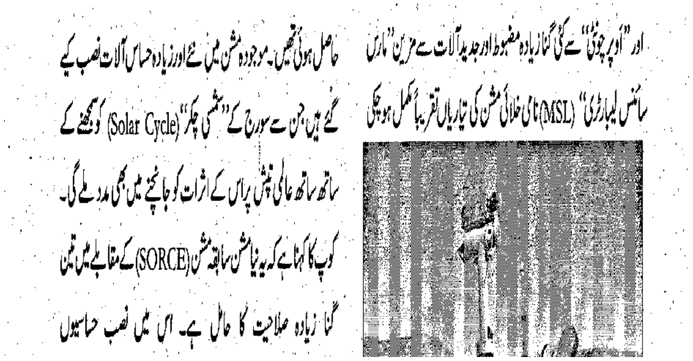


Page with single column



Page with two columns

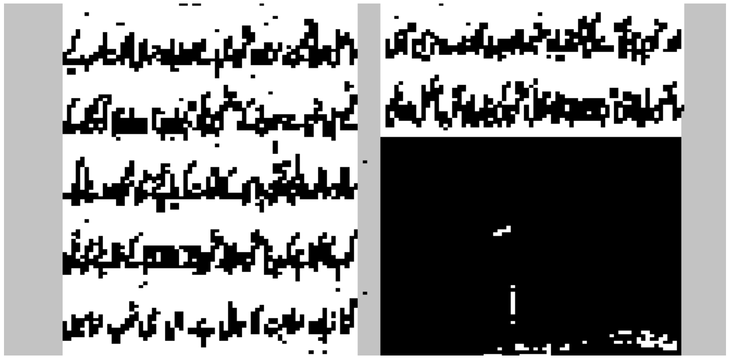
Layout Extraction



اور ''اوپر چوٹی'' سے کئی گنا زیادہ مضبوط اور جدید آلات سے مزین ''مارس
سائنس لیبارٹری'' (MSL) نامی خلائی مشن کی تیاریاں تقریباً مکمل ہو چکی

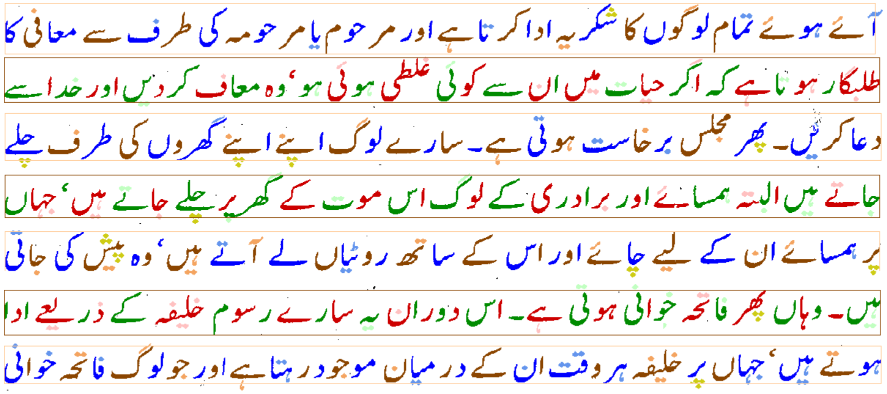
حاصل ہوئی تھیں۔ موجودہ مشن میں نئے اور زیادہ حساس آلات نصب کیے
گئے ہیں جن سے سورج کے ''شمسی چکر'' (Solar Cycle) کو سمجھنے کے
ساتھ ساتھ عالمی تپش پر اس کے اثرات کو جانچنے میں بھی مدد ملے گی۔
کوپ کا کہنا ہے کہ یہ نیا مشن سابقہ مشن (SORCE) کے مقابلے میں تین
گنا زیادہ صلاحیت کا حامل ہے۔ اس میں نصب حساسیوں

Page with two columns



White obstacle shown in grey color

Layout Extraction



آئے ہوئے تمام لوگوں کا شکریہ ادا کرتا ہے اور مرحوم یا مرحومہ کی طرف سے معافی کا
طلبگار ہوتا ہے کہ اگر حیات میں ان سے کوئی غلطی ہوئی ہو' وہ معاف کردیں اور خدا سے
دعا کریں۔ پھر مجلس برخاست ہوتی ہے۔ سارے لوگ اپنے اپنے گھروں کی طرف چلے
جاتے ہیں البتہ ہمسائے اور برادری کے لوگ اس موت کے گھر پر چلے جاتے ہیں' جہاں
پر ہمسائے ان کے لیے چائے اور اس کے ساتھ روٹیاں لے آتے ہیں' وہ پیش کی جاتی
ہیں۔ وہاں پھر فاتحہ خوانی ہوتی ہے۔ اس دوران یہ سارے رسوم خلیفہ کے ذریعے ادا
ہوتے ہیں' جہاں پر خلیفہ ہر وقت ان کے درمیان موجود رہتا ہے اور جو لوگ فاتحہ خوانی

Color coded output of a text area

106

آئے ہوئے تمام لوگوں کا شکریہ ادا کرتا ہے اور مرحوم یا مرحومہ کی طرف سے معافی کا
طلبگار ہوتا ہے کہ اگر حیات میں ان سے کوئی غلطی ہوئی ہو' وہ معاف کردیں اور خدا سے
دعا کریں۔ پھر مجلس برخاست ہوتی ہے۔ سارے لوگ اپنے اپنے گھروں کی طرف چلے
جاتے ہیں البتہ ہمسائے اور برادری کے لوگ اس موت کے گھر پر چلے جاتے ہیں' جہاں
پر ہمسائے ان کے لیے چائے اور اس کے ساتھ روٹیاں لے آتے ہیں' وہ پیش کی جاتی
ہیں۔ وہاں پھر فاتحہ خوانی ہوتی ہے۔ اس دوران یہ سارے رسوم خلیفہ کے ذریعے ادا
ہوتے ہیں' جہاں پر خلیفہ ہر وقت ان کے درمیان موجود رہتا ہے اور جو لوگ فاتحہ خوانی
کے لیے آتے ہیں ان کے لیے فاتحہ خوانی کراتا ہے۔

اس کے بعد تین دن تک اس گھر میں کچھ بھی نہیں پکتا بلکہ تین دن تک ہمسائے کے
لوگ' قبیلے کی خواتین اور رشتہ دار روزانہ تین اوقات میں چائے اور سالن و روٹیاں
لاتے ہیں۔ اس کے ساتھ ہی مرد لوگ بھی آس پاس سے اور دور دراز علاقوں سے
تعزیت کے واسطے آتے رہتے ہیں۔ پھر ان کے سامنے چائے و روٹی سے خاطر تواضع
بھی کی جاتی ہے۔ تین دن تک برادری کے لوگ بھی ساتھ رہتے ہیں اور سوگواروں
کی ہر طرح سے دلجوئی کرتے رہتے ہیں۔ تیسرے روز کی شام کو ایک بھیڑ وذبح کر کے
نذر و نیاز کا بندوبست کیا جاتا ہے جس میں برادری کے لوگوں کے علاوہ ہمسائے اور دیگر
رشتہ دار شامل ہوتے ہیں۔ اس رسم سوئم کی ادائیگی کے بعد بھی بعض خواتین اور مرد
جو اس دوران نہیں آسکے ہیں' ایک ہفتے تک تعزیت کے لیے آتے رہتے ہیں اور
مرحومین کے لیے فاتحہ خوانی کرتے ہیں۔

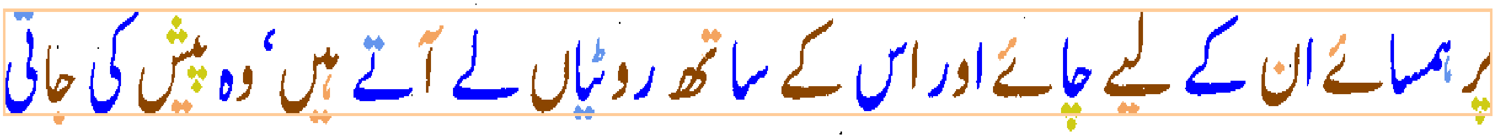
یہ بات بھی واضح رہے کہ اس رسم کی ادائیگی میں امیر و غریب کا کوئی امتیاز نہیں۔
سب کے لیے ایک قسم کی رسم ہے اور سب کی دلجوئی کی جاتی ہے۔ اس طرح غمزدہ
خاندان کو حوصلہ ملتا ہے اور ایک قسم کی بھائی چارگی کا احساس بڑھتا رہتا ہے۔ یہ روایت
اب تک برابر جاری ہے اور اس وقت ہنزہ والے خواہ پاکستان کے کسی بھی شہر میں رہتے
ہوں اسی رسم کی ادائیگی میں برابر کاربند ہیں اور اس طرح غمزدہ خاندان کو کہیں بھی بے
چارگی یا تنہائی کا احساس نہیں ہوتا۔

Page with single column

Accuracy – Text Extraction

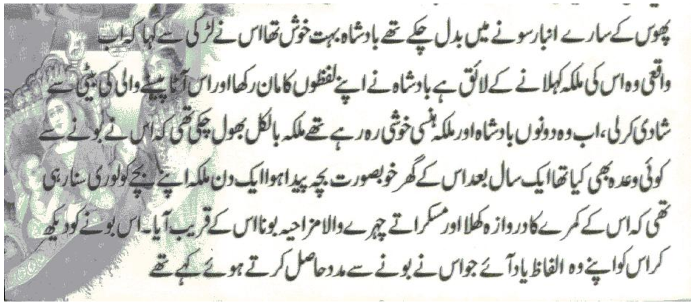
Total Images	1,257
Total Gold Textual CCs	2,092,782
Noise CCs	931,522
Gold Textual CCs excluding noise	1,161,260
Computed Textual CCs	1,140,959
Matched Textual CCs	1,140,712
Accuracy of Text as Text	98.23%

Accuracy – Line Segmentation

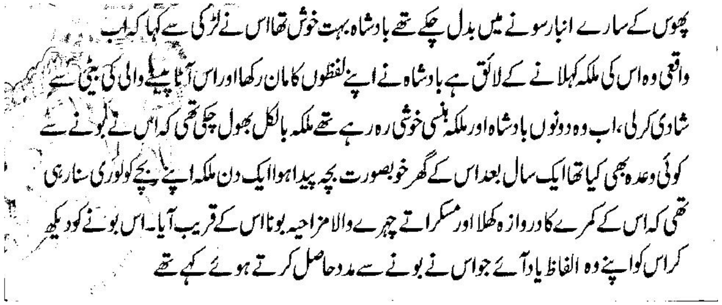


Font Size	Total Number of Lines	Correct Number of Lines	Accuracy
14	930	865	93.01%
16	820	758	92.44%
18	720	657	91.25%
20	570	439	77.02%
22	447	325	75.93%
24	700	659	94.14%
28	677	624	92.17%
32	775	679	87.61%
36	634	573	90.38%
40	519	443	85.36%

Layout Extraction Challenges


پھوس کے سارے انبار سونے میں بدل چکے تھے بادشاہ بہت خوش تھا اس نے لڑکی سے کہا کہ اب
واقعی وہ اس کی ملکہ کہلانے کے لائق ہے بادشاہ نے اپنے لفظوں کا مان رکھا اور اس آٹا پیسنے والی کی بیٹی سے
شادی کرلی، اب وہ دونوں بادشاہ اور ملکہ ہنسی خوشی رہ رہے تھے ملکہ بالکل بھول چکی تھی کہ اس نے بونے سے
کوئی وعدہ بھی کیا تھا ایک سال بعد اس کے گھر خوبصورت بچہ پیدا ہوا ایک دن ملکہ اپنے بچے کو لوری سنا رہی
تھی کہ اس کے کمرے کا دروازہ کھلا اور مسکراتے چہرے والا مزاحیہ بونا اس کے قریب آیا۔ اس بونے کو دیکھ
کر اس کو اپنے وہ الفاظ یاد آئے جو اس نے بونے سے مدد حاصل کرتے ہوئے کہے تھے

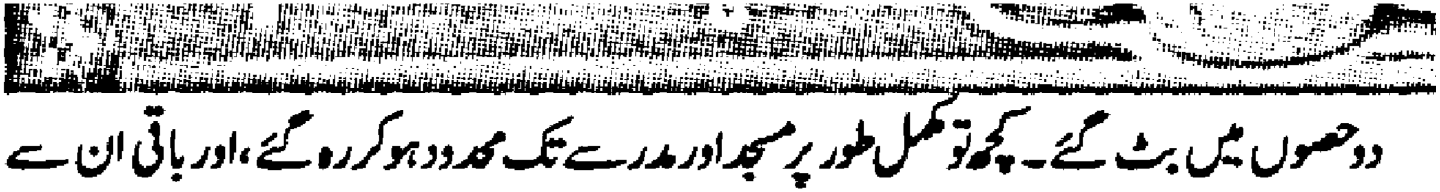
Half toner image of Font 22


پھوس کے سارے انبار سونے میں بدل چکے تھے بادشاہ بہت خوش تھا اس نے لڑکی سے کہا کہ اب
واقعی وہ اس کی ملکہ کہلانے کے لائق ہے بادشاہ نے اپنے لفظوں کا مان رکھا اور اس آٹا پیسنے والی کی بیٹی سے
شادی کرلی، اب وہ دونوں بادشاہ اور ملکہ ہنسی خوشی رہ رہے تھے ملکہ بالکل بھول چکی تھی کہ اس نے بونے سے
کوئی وعدہ بھی کیا تھا ایک سال بعد اس کے گھر خوبصورت بچہ پیدا ہوا ایک دن ملکہ اپنے بچے کو لوری سنا رہی
تھی کہ اس کے کمرے کا دروازہ کھلا اور مسکراتے چہرے والا مزاحیہ بونا اس کے قریب آیا۔ اس بونے کو دیکھ
کر اس کو اپنے وہ الفاظ یاد آئے جو اس نے بونے سے مدد حاصل کرتے ہوئے کہے تھے

Binarized version

Smeared version

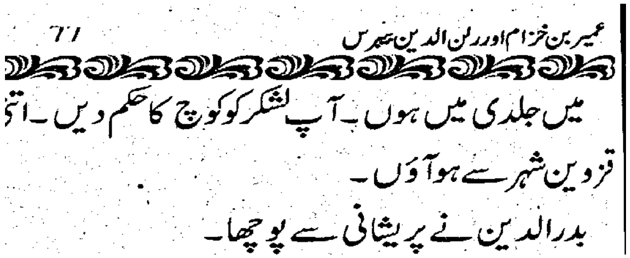
Layout Extraction Challenges



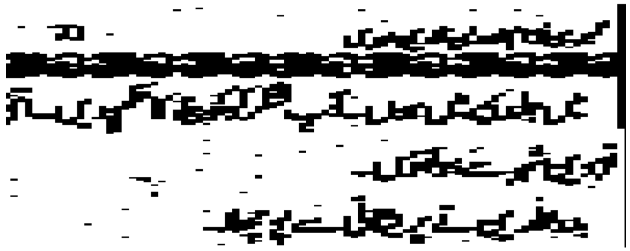
Main body joined with figure

Resultant smeared image

Layout Extraction Challenges



Page Frame in Binarized version



No White Obstacle detected

Lines merged with page frame and detected as figure