

# The Kiss of the Prince: Bringing Document Content to Life

Andreas Dengel



The idea for DAS was born on a bus ride in Las Vegas ...

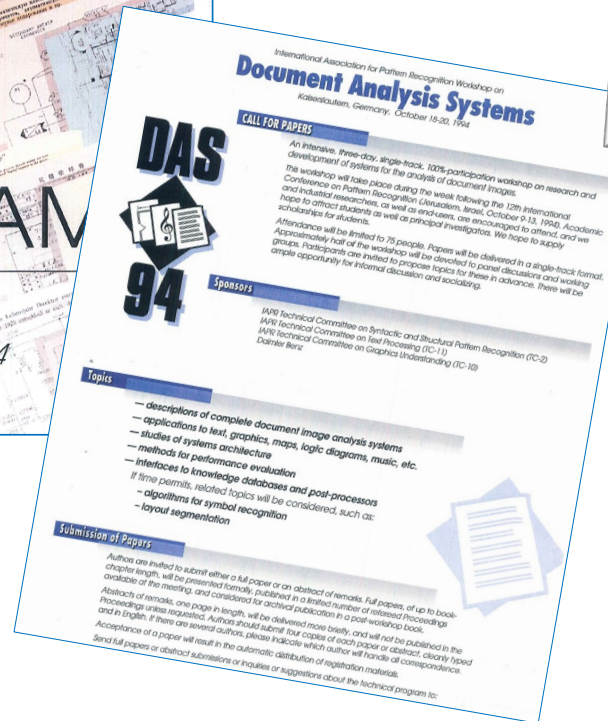
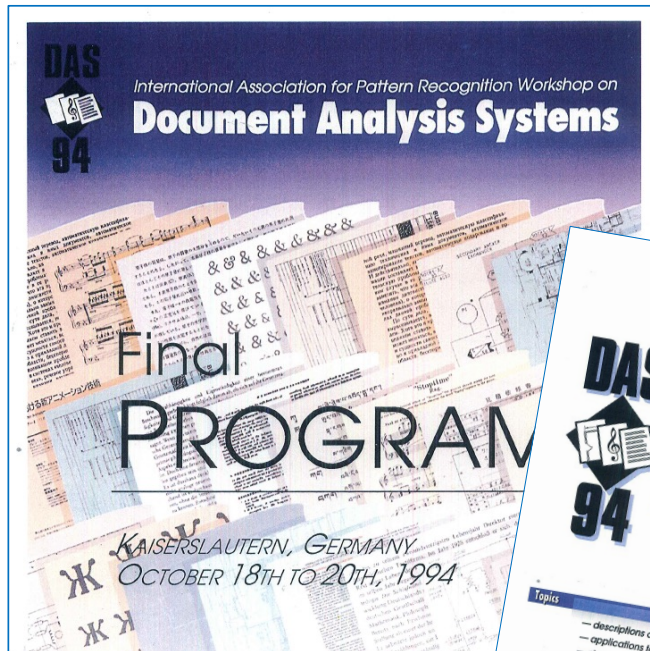


... aiming at justifying the growing demand on Document Analysis Solutions and Systems

*... in memory to my good old fellow Larry Spitz*



Supported by the IAPR Larry Spitz and I hosted a wonderful first DAS



# The start of the DAS workshop series was in 1994 in Kaiserslautern



*Andreas Dengel, A. Lawrence Spitz:*

Document Analysis Systems 1994, DFKI, Kaiserslautern, Germany, October 18-20, 1994, Proceedings

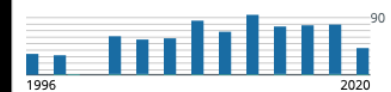
## DAS 1994

## Kaiserslautern



### [–] Venue statistics

#### records by year



#### document access ?

- closed (674)
- open (55)

#### frequent authors

- Andreas Dengel <sup>0001</sup> (31)
- Chew Lim Tan (24)
- Koichi Kise (23)
- Marcus Liwicki (20)
- Seiichi Uchida (20)
- Faisal Shafait (19)
- Jean-Marc Ogier (19)
- Basilios Gatos (19)
- C. V. Jawahar (19)
- Umapada Pal <sup>0001</sup> (18)

[show more](#)

<https://dblp.org/db/conf/das/index.html>

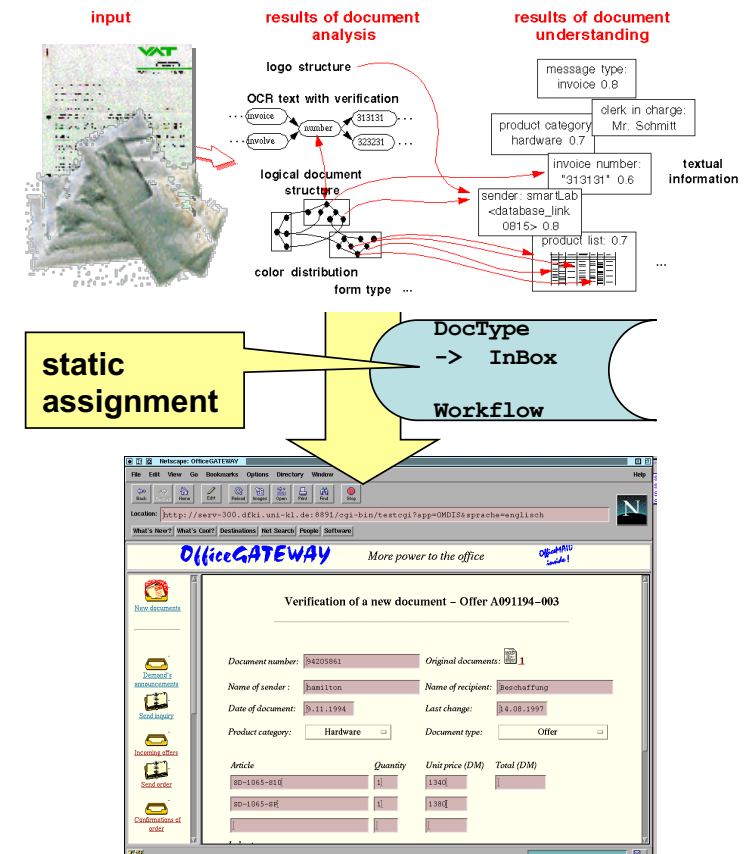


Image Source: <https://www.travellens.co/best-things-to-do-in-nagano/>

©andreas.dengel@dfki.de - 2022 - p. 4

# In 1994, we presented a first workflow-integrated solution OfficeMAID for purchase-to-pay processes

A. Dengel, R. Bleisinger, R. Hoch, F. Hönes, M. Malburg, and F. Fein, *OfficeMAID — A System for Automatic Mail Analysis, Interpretation and Delivery*, Proceedings DAS94, Kaiserslautern (Oct. 1994), pp. 253-276.



## OfficeMaid

- document analysis system with proprietary workflow
- after scanning and document analysis incoming (paper-) mail is embedded into purchase2pay process including a routing to the responsible clerk

## The second DAS location was Malvern in the beautiful Pennsylvania



*Jonathan J. Hull, Suzanne Liebowitz Taylor:*

Document Analysis Systems II - Second Workshop on Document Analysis Systems, DAS 1996, Malvern, PA, USA, October 14-16, 1996, Selected papers. Series in Machine Perception and Artificial Intelligence 29, WorldScientific 1998, ISBN 978-981-02-3103-3, pp. 1-528

# DAS 1996

## Malvern

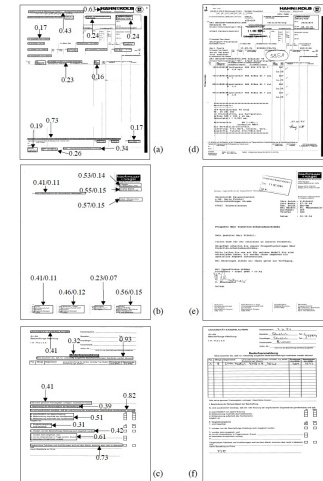
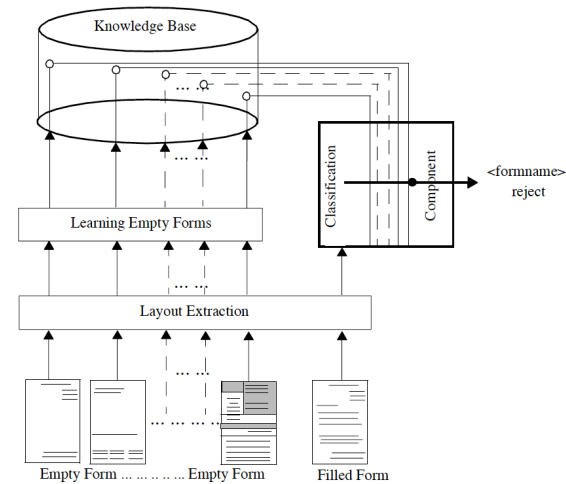


Image Source: [https://www.tripadvisor.de/Tourism-q53091-Malvern\\_Pennsylvania-Vacations.html](https://www.tripadvisor.de/Tourism-q53091-Malvern_Pennsylvania-Vacations.html)

©andreas.dengel@dfki.de - 2022 - p. 6

# Our focus at DAS 1996 was on FormClas, an OCR-free form classification system

F. Dubiel and A. Dengel, *FormClas — OCR-Free Classification of Forms*, Proceedings DAS96, Malvern, NJ, USA (Oct. 1996), pp. 40-60.



## FormClas ...

classified any type of filled form or document for which an empty reference pattern is known

In 1998, DAS took place at the former Olympics venue Nagano



*Seong-Whan Lee, Yasuaki Nakano:*

Document Analysis Systems: Theory and Practice, Third IAPR Workshop, DAS'98, Nagano, Japan, November 4-6, 1998, Selected Papers. Lecture Notes in Computer Science 1655, Springer 1999, ISBN 3-540-66507-2

**DAS 1998**

**Nagano**



Image Source: <https://www.travellens.co/best-things-to-do-in-nagano/>

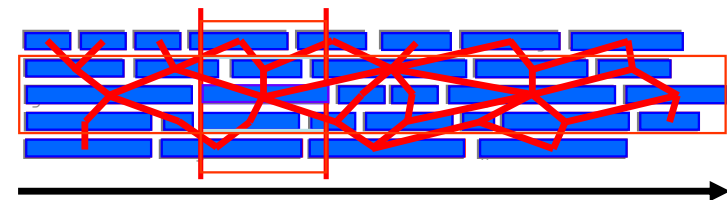
©andreas.dengel@dfki.de - 2022 - p. 8

# One of the papers we presented at DAS 1998, proposed a new clustering approach as part of the T-RECS system to find tables in printed documents

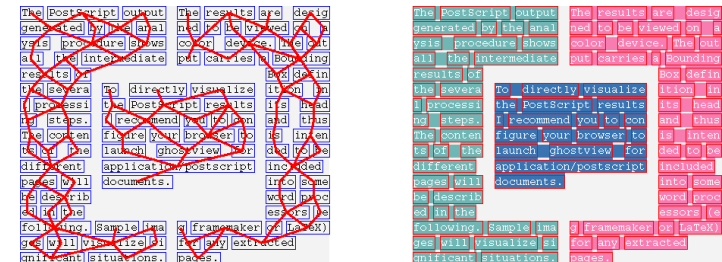
T. Kieninger and A. Dengel, *A Paper-to-HTML Table Converting System*,  
Proceedings DAS98, Nagano, Japan (Nov. 1998), pp. 356-365.

**T-RECS** follows a bottom-up clustering approach and was based on the following motivation:

- Large amounts of business letters contain the most important information within tables
- Insurance companies for example employ numerous clerks that enter data from paper documents into databases
- Due to the manyfold different senders and layouts, model based approaches are hardly applicable
- Model based approaches not only require strong efforts to set up the initial system, but also require permanent maintenance



- Definition of a window with horizontal and vertical segment overlaps
- Creation of a segmentation diagram



## Table-Location:

T-Recs++

Determined  
blocksegments

| Elektronix  |           | Auftragsbestätigung   |      |                             |  |                     |                |                |
|---|-----------|---|------|-----------------------------|--|---------------------|----------------|----------------|
| Datum: 07.11.94<br>Blm: 001<br>Versandt: ELAN/RINDT FREI<br>Messung/Gesamt: 45163-001<br>AG SYSTEMS/WARE<br>GEBÄUDE 36<br>PAUL-EHRICH-STR.<br>67663 KAISERSLAUTERN<br>Bei Schriftwechsel oder Zahlung angeben:<br>X Kurdes-Nr: 45163-001<br>X Auftrags-Nr: 170681-23681<br>Rechnungs-Nr:<br>Buchstabs-Nr:   |           | Tektronix GmbH, Deutschschreibstraße, Gülden Allee 13, 51047 Köln<br>Postfach 100634, 50455 Köln<br>UNIVERSITÄT KAISERSLAUTERN<br>BESCHAFFUNGSREFERAT<br>ERWIN-SCHROEDINGER-STR.<br>67663 KAISERSLAUTERN<br>Ihre Zeichen: 896800/140/94 |      |                             |  |                     |                |                |
| Pos.-Nr.  | Sech.-Nr. | Type  | Qtd. | Abf.                        | Beschreibung                                 | Menge               | Einzelpreis DM | Gesamtpreis DM |
| Wir bedanken uns für Ihren Auftrag und bestätigen ihn wie folgt:  |           |   |      |                             |  |                     |                |                |
| 001   | XP354     |   |      |                             | 19 ZOLL GRAUSTUFEN X TERMINAL<br>ECCN: 4A966 | 1                   | 5000.00        |                |
|   |           | A1  |      |                             | MIT DEUTSCHEM NETZKABEL                      |                     | 0.00           |                |
|   |           | B9  |      |                             | X SERVER BENUTZUNGS-LIZENZ                   |                     | 0.00           |                |
|   |           |   |      |                             | POSITIONS-ZWISCHENSUMME                      |                     |                | 5000.00        |
|   |           |   |      |                             | Vorauss. Lieferdat. 19.11.94                 |                     |                |                |
| 002   | XPFXG     |   |      |                             | IBM 102 TASTATUR (DEUTSCH)<br>ECCN: 4A966    | 1                   | 0.00           |                |
|   |           |   |      |                             | Vorauss. Lieferdat. 19.11.94                 |                     |                |                |
| 003   | XP354     |   |      |                             | 19 ZOLL GRAUSTUFEN X TERMINAL<br>ECCN: 4A966 | 1                   | 5000.00        |                |
|   |           | A1  |      |                             | MIT DEUTSCHEM NETZKABEL                      |                     | 0.00           |                |
|   |           | B9  |      |                             | X SERVER BENUTZUNGS-LIZENZ                   |                     | 0.00           |                |
|   |           |   |      |                             | POSITIONS-ZWISCHENSUMME                      |                     |                | 5000.00        |
|   |           |   |      |                             | Vorauss. Lieferdat. 19.11.94                 |                     |                |                |
| 004   | XPFXH     |   |      |                             | IBM 101 TASTATUR (US)<br>ECCN: 4A966         | 1                   | 0.00           |                |
|   |           |   |      |                             | Vorauss. Lieferdat. 19.11.94                 |                     |                |                |
| Bitte wenden Sie sich bei Rückfragen an die unten genannte Adresse.   |           |   |      |                             |  |                     |                |                |
| Mit freundlichen Grüßen   |           |   |      |                             |  |                     |                |                |
| Ihre<br>TEKTRONIX GmbH  |           |   |      |                             |  |                     |                |                |
| Der Export/Re-Export bestimmter Elektronix-Produkte unterliegt bestimmten technischen Bedingungen und ist durch das Export Control System (ECCN) geregelt. Die Kunden von Elektronix sind für die Einhaltung dieser Bestimmungen verantwortlich und sind verpflichtet, sich vor einem Export bei Elektronix über Exportbeschränkungen zu informieren.<br>Elektronix ist eine Tochtergesellschaft von Tektronix, Inc. (USA).<br>Elektronix ist eine Tochtergesellschaft von Tektronix, Inc. (USA).<br>Elektronix ist eine Tochtergesellschaft von Tektronix, Inc. (USA).<br>Elektronix ist eine Tochtergesellschaft von Tektronix, Inc. (USA). |           |   |      |                             |  |                     |                |                |
| Nettopreis DM   |           | MwSt. 19%   |      | MwSt. 19%                   |  | Rechnungsnummer DM  |                |                |
| 10000.00  |           | zzgl. ges. UST  |      | 30 TAGE NETTO               |  |                     |                |                |
| Bei Rückfragen  |           | Karl-Georg Meerkamp   |      | 0221/96969-361              |  | Telefax: /96969-362 |                |                |
| Commerzbank AG, Postfach 100634, 50455 Köln   |           | Postfach 100634, 50455 Köln   |      | Postfach 100634, 50455 Köln |  |                     |                |                |

## Table-Location:

T-Recs++

Determination  
of column-  
fragments using  
the  
similarity  
relation  
 $\text{sim}_{\epsilon}(a,b)$

| Tektronix   |           | Auftragsbestätigung  |      |      |                               |       |                |                |
|---|-----------|--|------|------|-------------------------------|-------|----------------|----------------|
| Datum: 07.11.94   |           | Hin: 001   |      |      |                               |       |                |                |
| Versandt: ELAN/RINDT FREI   |           | Wegschreiben   |      |      |                               |       |                |                |
| Wegschreiben  |           | 45163-001  |      |      |                               |       |                |                |
| AG SYSTEMS/WARE   |           | GEBÄUDE 36   |      |      |                               |       |                |                |
| PAUL-EHRICH-STR.  |           | 67663 KAISERSLAUTERN   |      |      |                               |       |                |                |
| Bei Schriftwechsel oder Zahlung angeben   |           | <input checked="" type="checkbox"/> Kurzlos-Nr. 45163-001<br><input checked="" type="checkbox"/> Auftrags-Nr. 170681-23681<br><input type="checkbox"/> Rechnungs-Nr.<br><input type="checkbox"/> Buchst.-Nr. |      |      |                               |       |                |                |
| Ihre Zeichen: 896800/140/94   |           |  |      |      |                               |       |                |                |
| Pos.-Nr.  | Sech.-Nr. | Typ  | Qtd. | Abf. | Bezeichnung                   | Menge | Einzelpreis DM | Gesamtpreis DM |
| Wir bedanken uns für Ihren Auftrag und bestätigen ihn wie folgt:  |           |  |      |      |                               |       |                |                |
|   |           |  |      |      | 19 ZOLL GRAUSTUFEN X TERMINAL |       |                |                |
|   |           |  |      |      | ECCN: 4A96G                   |       |                |                |
|   |           |  |      |      | MIT DEUTSCHEM NETZKABEL       |       |                |                |
|   |           |  |      |      | X SERVER BENUTZUNGS-LIZENZ    |       |                |                |
|   |           |  |      |      | POSITIONS-ZWISCHENSUMME       |       |                |                |
|   |           |  |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
|   |           |  |      |      | IBM 102 TASTATUR (DEUTSCH)    |       |                |                |
|   |           |  |      |      | ECCN: 4A96G                   |       |                |                |
|   |           |  |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
|   |           |  |      |      | 19 ZOLL GRAUSTUFEN X TERMINAL |       |                |                |
|   |           |  |      |      | ECCN: 4A96G                   |       |                |                |
|   |           |  |      |      | MIT DEUTSCHEM NETZKABEL       |       |                |                |
|   |           |  |      |      | X SERVER BENUTZUNGS-LIZENZ    |       |                |                |
|   |           |  |      |      | POSITIONS-ZWISCHENSUMME       |       |                |                |
|   |           |  |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
|   |           |  |      |      | IBM 101 TASTATUR (US)         |       |                |                |
|   |           |  |      |      | ECCN: 4A96G                   |       |                |                |
|   |           |  |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
| Bitte wenden Sie sich bei Rückfragen an die unten genannte Adresse.   |           |  |      |      |                               |       |                |                |
| Mit freundlichen Grüßen   |           |  |      |      |                               |       |                |                |
| Ihre<br>TEKTRONIX GmbH  |           |  |      |      |                               |       |                |                |
| Der Export/Re-Export bestimmter Tektronix-Produkte unterliegt bestimmten technischen Bedingungen. Tektronix-Produkte sind gekennzeichnet mit dem Export Control Code (ECCN), der das Exportverhalten gemäß der Außenwirtschaftsverordnung festlegt. Alle Kunden von Tektronix sind für die Einhaltung dieser Bestimmungen verantwortlich und sind verpflichtet, sich vor einem Export bei Tektronix über Exportbeschränkungen zu informieren. |           |  |      |      |                               |       |                |                |
| Bei Rückfragen: Karl-Georg Meerkamp   |           |  |      |      |                               |       |                |                |
| 0000.00 zzgl. ges. UST  |           |  |      |      |                               |       |                |                |
| 30 TAGE NETTO   |           |  |      |      |                               |       |                |                |
| 0221/96969-361 / 96969-362  |           |  |      |      |                               |       |                |                |

## Table-Location:

T-Recs++

Projection of  
the column-  
fragments onto  
the y-axis to  
evaluate the  
vertical  
expansion of  
the table

y-axis

| Tektronix   |          | Auftragsbestätigung   |      |      |  |       |                |                |
|---|----------|---|------|------|--|-------|----------------|----------------|
| Datum: 07.11.94   |          | Hin: 001  |      |      |  |       |                |                |
| Telefon: 0611 40044-11, 40044-12, 40044-13, 40044-14, 40044-15, 40044-16, 40044-17, 40044-18, 40044-19, 40044-20, 40044-21, 40044-22, 40044-23, 40044-24, 40044-25, 40044-26, 40044-27, 40044-28, 40044-29, 40044-30, 40044-31, 40044-32, 40044-33, 40044-34, 40044-35, 40044-36, 40044-37, 40044-38, 40044-39, 40044-40, 40044-41, 40044-42, 40044-43, 40044-44, 40044-45, 40044-46, 40044-47, 40044-48, 40044-49, 40044-50, 40044-51, 40044-52, 40044-53, 40044-54, 40044-55, 40044-56, 40044-57, 40044-58, 40044-59, 40044-60, 40044-61, 40044-62, 40044-63, 40044-64, 40044-65, 40044-66, 40044-67, 40044-68, 40044-69, 40044-70, 40044-71, 40044-72, 40044-73, 40044-74, 40044-75, 40044-76, 40044-77, 40044-78, 40044-79, 40044-80, 40044-81, 40044-82, 40044-83, 40044-84, 40044-85, 40044-86, 40044-87, 40044-88, 40044-89, 40044-90, 40044-91, 40044-92, 40044-93, 40044-94, 40044-95, 40044-96, 40044-97, 40044-98, 40044-99, 40044-100 |          | Versandart: ELAN/RINDT FREI   |      |      |  |       |                |                |
| Telefon: 0611 40044-11, 40044-12, 40044-13, 40044-14, 40044-15, 40044-16, 40044-17, 40044-18, 40044-19, 40044-20, 40044-21, 40044-22, 40044-23, 40044-24, 40044-25, 40044-26, 40044-27, 40044-28, 40044-29, 40044-30, 40044-31, 40044-32, 40044-33, 40044-34, 40044-35, 40044-36, 40044-37, 40044-38, 40044-39, 40044-40, 40044-41, 40044-42, 40044-43, 40044-44, 40044-45, 40044-46, 40044-47, 40044-48, 40044-49, 40044-50, 40044-51, 40044-52, 40044-53, 40044-54, 40044-55, 40044-56, 40044-57, 40044-58, 40044-59, 40044-60, 40044-61, 40044-62, 40044-63, 40044-64, 40044-65, 40044-66, 40044-67, 40044-68, 40044-69, 40044-70, 40044-71, 40044-72, 40044-73, 40044-74, 40044-75, 40044-76, 40044-77, 40044-78, 40044-79, 40044-80, 40044-81, 40044-82, 40044-83, 40044-84, 40044-85, 40044-86, 40044-87, 40044-88, 40044-89, 40044-90, 40044-91, 40044-92, 40044-93, 40044-94, 40044-95, 40044-96, 40044-97, 40044-98, 40044-99, 40044-100 |          | Vorgeschrieben: 45163-001   |      |      |  |       |                |                |
| UNIVERSITÄT KAISERSLAUTERN<br>BESCHAFFUNGSREFERAT   |          | AG SYSTEMS/WARE<br>GEBÄUDE 36   |      |      |  |       |                |                |
| ERWIN-SCHROEDINGER-STR.   |          | PAUL-EHRICH-STR.  |      |      |  |       |                |                |
| 67663 KAISERSLAUTERN  |          | 67663 KAISERSLAUTERN  |      |      |  |       |                |                |
| Ihre Zeichen: 896800/140/94   |          | Bei Schriftwechsel oder Zahlung angeben:<br>X Kurkopf-Nr: 45163-001<br>X Auftrags-Nr: 170681-23681<br>Rechungs-Nr:<br>Zustell-Nr: |      |      |  |       |                |                |
| Pos.-Nr.  | Bech-Nr. | Type  | Qtd. | Unit | Bezeichnung                                  | Menge | Einzelpreis DM | Gesamtpreis DM |
| Wir bedanken uns für Ihren Auftrag und bestätigen ihn wie folgt:  |          |   |      |      |  |       |                |                |
|   |          |   |      |      | 19 ZOLL GRAUSTUFEN X TERMINAL<br>ECCN: 4A96G |       |                |                |
|   |          |   |      |      | MIT DEUTSCHEM NETZKABEL                      |       |                |                |
|   |          |   |      |      | X SERVER BENUTZUNGS-LIZENZ                   |       |                |                |
|   |          |   |      |      | POSITIONS-ZWISCHENSUMME                      |       |                |                |
|   |          |   |      |      | Vorauss. Lieferdat. 19.11.94                 |       |                |                |
|   |          |   |      |      | IBM 102 TASTATUR (DEUTSCH)                   |       |                |                |
|   |          |   |      |      | ECCN: 4A96G                                  |       |                |                |
|   |          |   |      |      | Vorauss. Lieferdat. 19.11.94                 |       |                |                |
|   |          |   |      |      | 19 ZOLL GRAUSTUFEN X TERMINAL<br>ECCN: 4A96G |       |                |                |
|   |          |   |      |      | MIT DEUTSCHEM NETZKABEL                      |       |                |                |
|   |          |   |      |      | X SERVER BENUTZUNGS-LIZENZ                   |       |                |                |
|   |          |   |      |      | POSITIONS-ZWISCHENSUMME                      |       |                |                |
|   |          |   |      |      | Vorauss. Lieferdat. 19.11.94                 |       |                |                |
|   |          |   |      |      | IBM 101 TASTATUR (US)                        |       |                |                |
|   |          |   |      |      | ECCN: 4A96G                                  |       |                |                |
|   |          |   |      |      | Vorauss. Lieferdat. 19.11.94                 |       |                |                |
| Bitte wenden Sie sich bei Rückfragen an die unten<br>genannte Adresse.  |          |   |      |      |  |       |                |                |
| Mit freundlichen Grüßen   |          |   |      |      |  |       |                |                |
| Ihre<br>TEKTRONIX GmbH  |          |   |      |      |  |       |                |                |
| Der Export/Re-Export bestimmter Tektronix-Produkte ist ohne technische<br>Daten untersagt. Bitte beachten Sie die Exportkontrollbestimmungen.<br>Rechtliche Produkte sind gekennzeichnet mit dem Export Control<br>Number (ECN), die das Kennzeichen gemäß der Außen-<br>wirtschaftsverordnung enthält.<br>Alle Kunden von Tektronix sind für die Einhaltung dieser Bestimmungen<br>verantwortlich und sind verpflichtet, sich vor einem Export bei Tektronix über<br>Exportbeschränkungen zu informieren.  |          |   |      |      |  |       |                |                |
| Es gelten unsere Verkaufs- und Lieferbedingungen (siehe Informationsmaterial) für<br>Kundenbeziehungen. Keine Gewähr für die Richtigkeit der Angaben.   |          |   |      |      |  |       |                |                |
| Betreiber: Tektronix GmbH   |          |   |      |      |  |       |                |                |
| Spezialbedingung: Keine Zahlung   |          |   |      |      |  |       |                |                |
| Rechnungsdatum: 07.11.94  |          |   |      |      |  |       |                |                |
| Commerzbank AG, Postfach 100 000, 60000 Frankfurt   |          |   |      |      |  |       |                |                |
| Hilfsweise DM   |          |   |      |      |  |       |                |                |
| MwSt. Nr.   |          |   |      |      |  |       |                |                |
| MwSt. 10%   |          |   |      |      |  |       |                |                |
| Rechnungsnummer DM  |          |   |      |      |  |       |                |                |
| 0000.00   |          |   |      |      |  |       |                |                |
| zzgl. ges. UST  |          |   |      |      |  |       |                |                |
| Zahlungsbefristung  |          |   |      |      |  |       |                |                |
| 30 TAGE NETTO   |          |   |      |      |  |       |                |                |
| Bei Rückfragen: Karl-Georg Meerkamp   |          |   |      |      |  |       |                |                |
| 0221/96969-361  |          |   |      |      |  |       |                |                |
| Telefax: /96969-362   |          |   |      |      |  |       |                |                |

## Table-Location:

T-Recs++

## Resulting table region

**Tektronix**

**Auftragsbestätigung**

**Datum:** 07.11.94

**Hin:** 001

**Versandt:** ELAN/RINDT FREI

**Wegung:**

**Wegungsschein:** 45163-001

**Tektronix GmbH, Deutschschreibtele, Gelsen-Allee 13, 41007 RSP**

**Postfach:** 100634, 50475 Köln

**UNIVERSITÄT KAISERSLAUTERN**

**BESCHAFFUNGSREFERAT**

**ERWIN-SCHROEDINGER-STR.**

**67663 KAISERSLAUTERN**

**AG SYSTEMS/WARE**

**GEBAUDE 36**

**PAUL-EHRICH-STR.**

**67663 KAISERSLAUTERN**

**Bei Schriftwechsel oder Zahlung angeben:**

☒
**Kontop.-Nr.** 45163-001

☒
**Auftrags-Nr.** 170681-23681

☐
**Rechnungs-Nr.**

☐
**Einzugs-Nr.**

**Ihre Zeichen:** 896800/140/94

| Pos.-Nr.   | Bech-Nr. | Type | Qtd. | Abf. | Bezeichnung                   | Menge | Einzelpreis DM | Gesamtpreis DM |
|--|----------|------|------|------|-------------------------------|-------|----------------|----------------|
| Wir bedanken uns für Ihren Auftrag und bestätigen ihn wie folgt: |          |      |      |      |                               |       |                |                |
|  |          |      |      |      | 19 ZOLL GRAUSTUFEN X TERMINAL |       |                |                |
|  |          |      |      |      | ECCN: 4A966                   |       |                |                |
|  |          |      |      |      | MIT DEUTSCHEM NETZKABEL       |       |                |                |
|  |          |      |      |      | X SERVER BENUTZUNGS-LIZENZ    |       |                |                |
|  |          |      |      |      | POSITIONS-ZWISCHENSUMME       |       |                |                |
|  |          |      |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
|  |          |      |      |      | IBM 102 TASTATUR (DEUTSCH)    |       |                |                |
|  |          |      |      |      | ECCN: 4A966                   |       |                |                |
|  |          |      |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
|  |          |      |      |      | 19 ZOLL GRAUSTUFEN X TERMINAL |       |                |                |
|  |          |      |      |      | ECCN: 4A966                   |       |                |                |
|  |          |      |      |      | MIT DEUTSCHEM NETZKABEL       |       |                |                |
|  |          |      |      |      | X SERVER BENUTZUNGS-LIZENZ    |       |                |                |
|  |          |      |      |      | POSITIONS-ZWISCHENSUMME       |       |                |                |
|  |          |      |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |
|  |          |      |      |      | IBM 101 TASTATUR (US)         |       |                |                |
|  |          |      |      |      | ECCN: 4A966                   |       |                |                |
|  |          |      |      |      | Vorauss. Lieferdat. 19.11.94  |       |                |                |

Bitte wenden Sie sich bei Rückfragen an die unten genannte Adresse.

Mit freundlichen Grüßen

Ihre  
TEKTRONIX GmbH

**Der Export/Re-Export bestimmter Tektronix Produkte unter technischer**

**Daten unterliegt den deutschen (in GOST) bzw. den US-Exportkontroll-**

**Bestimmungen.**

**Rechtliche Produkte sind gekennzeichnet mit dem Export Commodity**

**Control Number (ECCN), die das Kennzeichen gemäß der Außen-**

**wirtschaftsverordnung enthält.**

**Alle Kunden von Tektronix sind für die Einhaltung dieser Bestimmungen**

**verantwortlich und sind gehalten, sich vor einem Export bei Tektronix über**

**Exportklassen zu informieren.**

**Es gelten unsere Verkaufs- und Gewinnaussagen (Preis-Mitteilungsgemeinschaft) für**

**Kundenbeziehungen, die in einem Jahr abgeschlossen werden.**

**Unterstützung:** 100-10000

**Copyright:** 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024

**Commerzbank AG, Postfach 100 100, 50475 Köln**

**Netto:** 10000,00

**MwSt:** 10%

**MwSt:** 10%

**Rechnungsnetto:** 10000,00

**Zahlungsbefristung:** 30 TAGE NETTO

**Bei Rückfragen:** Karl-Georg Meerkamp

**Telefon:** 0221/96969-361

**Telefax:** 0221/96969-362

## T-Recs++

## Integration of other blocks that are covered by the rectangular table-area

## Determination of the row- and column structure

## Elektronik

Telefon: 09401, Deutsch: 09401-10000, 09401-10001  
 Telefax: 09401, Deutsch: 09401-10000, 09401-10001  
 Telefax: 09401, Deutsch: 09401-10000, 09401-10001

## Auftragsbestätigung

Datum: 07.11.94  
 Uhrzeit: 00:01

UNIVERSITÄT KAISERSLAUTERN  
 BESCHAFFUNGSREFERAT

ERWIN-SCHROEDINGER-STR.

67663 KAISERSLAUTERN

AG SYSTEMSOFTWARE  
 GEBÄUDE 36

PAUL-EHRICH-STR.

67663 KAISERSLAUTERN

Ihre Zeichen: 896800/140/94

Bei Bedarf:

X Kunden-Nr.  
 X Auftrags-Nr.  
 X Rechnungs-Nr.  
 X Durchschnitt-Nr.

| Pos.-Nr.   | Batch-Nr. | Type | Qtd. | Mod. | Bezeichnung | Menge | Einzelpreis | Summe |
|--|-----------|------|------|------|-------------|-------|-------------|-------|
| Wir bedanken uns für Ihren Auftrag und bestätigen ihn wie folgt: |           |      |      |      |             |       |             |       |
| 1  |           |      |      |      |             |       |             |       |
| 2  |           |      |      |      |             |       |             |       |
| 3  |           |      |      |      |             |       |             |       |
| 4  |           |      |      |      |             |       |             |       |
| 5  |           |      |      |      |             |       |             |       |
| 6  |           |      |      |      |             |       |             |       |
| 7  |           |      |      |      |             |       |             |       |
| 8  |           |      |      |      |             |       |             |       |
| 9  |           |      |      |      |             |       |             |       |
| 10   |           |      |      |      |             |       |             |       |
| 11   |           |      |      |      |             |       |             |       |
| 12   |           |      |      |      |             |       |             |       |
| 13   |           |      |      |      |             |       |             |       |
| 14   |           |      |      |      |             |       |             |       |
| 15   |           |      |      |      |             |       |             |       |
| 16   |           |      |      |      |             |       |             |       |
| 17   |           |      |      |      |             |       |             |       |
| 18   |           |      |      |      |             |       |             |       |
| 19   |           |      |      |      |             |       |             |       |
| 20   |           |      |      |      |             |       |             |       |
| 21   |           |      |      |      |             |       |             |       |
| 22   |           |      |      |      |             |       |             |       |
| 23   |           |      |      |      |             |       |             |       |
| 24   |           |      |      |      |             |       |             |       |
| 25   |           |      |      |      |             |       |             |       |
| 26   |           |      |      |      |             |       |             |       |
| 27   |           |      |      |      |             |       |             |       |
| 28   |           |      |      |      |             |       |             |       |
| 29   |           |      |      |      |             |       |             |       |
| 30   |           |      |      |      |             |       |             |       |

Bitte wenden Sie sich bei Rückfragen an die unten genannte Adresse.

Mit freundlichen Grüßen

Ihre  
 ELEKTRONIK GmbH

Der Export/Re-Export bestimmter Elektronik-Produkte sowie technischer Daten unterliegt den Vorschriften (in CoCo) bzw. den US-Export-Kontrollbestimmungen (ECCN), die den Re-Export gemäß Art. 1 der Außenwirtschaftsverordnung einschränken.

Alle Kunden von ELEKTRONIK sind für die Einhaltung dieser Bestimmungen verantwortlich und garantieren, dass vor einem Export bei ELEKTRONIK über Exportkontrollen als Informanten.

Hilfsweise DM: 10000,00

Zahlungsbefreiung: 30 TAGE NETTO

Rechnungsnummer: 10000,00

zzgl. ges. UST

Bei Rückfragen: Karl-Georg Meenken

Telefon: 0221/96969-361

Telefax: 0221/96969-362

## Table-Location:

T-Recs++

Evaluation of an overall evidence for all potential tables based on the similarity within the column fragments

Remove table hypothesis with too little evidence

**Tektronix**

Tektronix GmbH, Deutschschreibtele, Gellena Allee 13, 61007 Bsp  
Postfach  
Postfach 100, 61007 Bsp

UNIVERSITÄT KAISERSLAUTERN  
BESCHAFFUNGSREFERAT

ERWIN-SCHROEDINGER-STR.

67663 KAISERSLAUTERN

Ihre Zeichen: 896800/140/94

**Auftragsbestätigung**

Datum: 07.11.94 Hin: 001

Versandt: ELAN/RINDT FREI  
Nebenkonto:

Abrechnung: 45163-001

AG SYSTEMSOFWARE  
GEBÄUDE 30

PAUL-EHRICH-STR.

67663 KAISERSLAUTERN

Bei Schriftwechsel oder Zahlung angeben:

X Kunden-Nr. 45163-001  
X Auftrags-Nr. 170681-23681

Rechnungs-Nr.  
Zustell-Nr.

| Pos.-Nr.   | Seit-Nr. | Type | Qtd. | Art | Bezeichnung | Menge | Einzelpreis DM | Gesamtpreis DM |
|--|----------|------|------|-----|-------------|-------|----------------|----------------|
| Wir bedanken uns für Ihren Auftrag und bestätigen ihn wie folgt: |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |
|  |          |      |      |     |             |       |                |                |

Bitte wenden Sie sich bei Rückfragen an die unten genannte Adresse.

Mit freundlichen Grüßen

Ihre  
TEKTRONIX GmbH

Der Export/Re-Export bestimmter Tektronix Produkte ist ohne technische Daten untersagt. Bitte beachten Sie die Exportkontrollbestimmungen.  
Rechtliche Produkte sind gekennzeichnet mit dem Export Control Number (ECN), die das Ministerium gemäß der Außenwirtschaftsverordnung erteilt.  
Alle Kunden von Tektronix sind für die Einhaltung dieser Bestimmungen verantwortlich und sind verpflichtet, sich vor einem Export bei Tektronix über Exportlizenzen zu informieren.

Es gelten unsere Verkaufs- und allgemeine Geschäftsbedingungen (AGB) und unsere Lieferbedingungen (LVB) sowie unsere AGB zur Garantie.

Bank für den Export: KfW-Bank

Bank für den Import: KfW-Bank

Bank für den Export: KfW-Bank

Bank für den Import: KfW-Bank

Netto: 10000,00 zzgl. ges. UST

Zahlungsschuldigkeit: 30 TAGE NETTO

Bei Rückfragen: Karl-Georg Meerkamp

Telefon: 0221/96969-361

Fax: 0221/96969-362

In 1998, I founded Insiders, the first start-up company of DFKI with five researchers of my group, which rapidly developed and today has about 1500 customers worldwide



For more than two years, I also acted as the CEO of Insiders and reduced my job as a professor to 50%



Insiders was established about 20 years ago and is the most successful spin-off of the German Research Center for Artificial Intelligence (DFKI). As an award-winning IT company, Insiders Technologies has offices in Kaiserslautern, renowned for being a high-tech region, and in Berlin, the stronghold of start-ups. It has long since become a leader in technology for the digitalization of business processes. The most up-to-date methods of Artificial Intelligence are the basis for the success of Insiders' highly efficient products.

insiders  
technologies

Company

**FROM SPIN-OFF TO GLOBAL PLAYER**



DAS 2000 took place in Southern America



***NO PROCEEDING AVAILABLE!***

**DAS 2000**

**Rio de Janeiro**



In 2000, had really trouble because the original plan was to cancel DAS!



S. Klink, A. Dengel, and T. Kieninger, *Document Structure Analysis Based on Layout and Textual Features*, Proceedings DAS2000, Rio de Janeiro, Brazil (Dec. 2000), pp. 41-52.

Nothing special to report except  
that the organization was so bad  
that I didn't show up in Rio!

## In 2002, DAS was hosted at the campus of Princeton University



*Daniel P. Lopresti, Jianying Hu, Ramanujan S. Kashi:*

Document Analysis Systems V, 5th International Workshop, DAS 2002, Princeton, NJ, USA, August 19-21, 2002, Proceedings. Lecture Notes in Computer Science 2423, Springer 2002, ISBN 3-540-44068-2

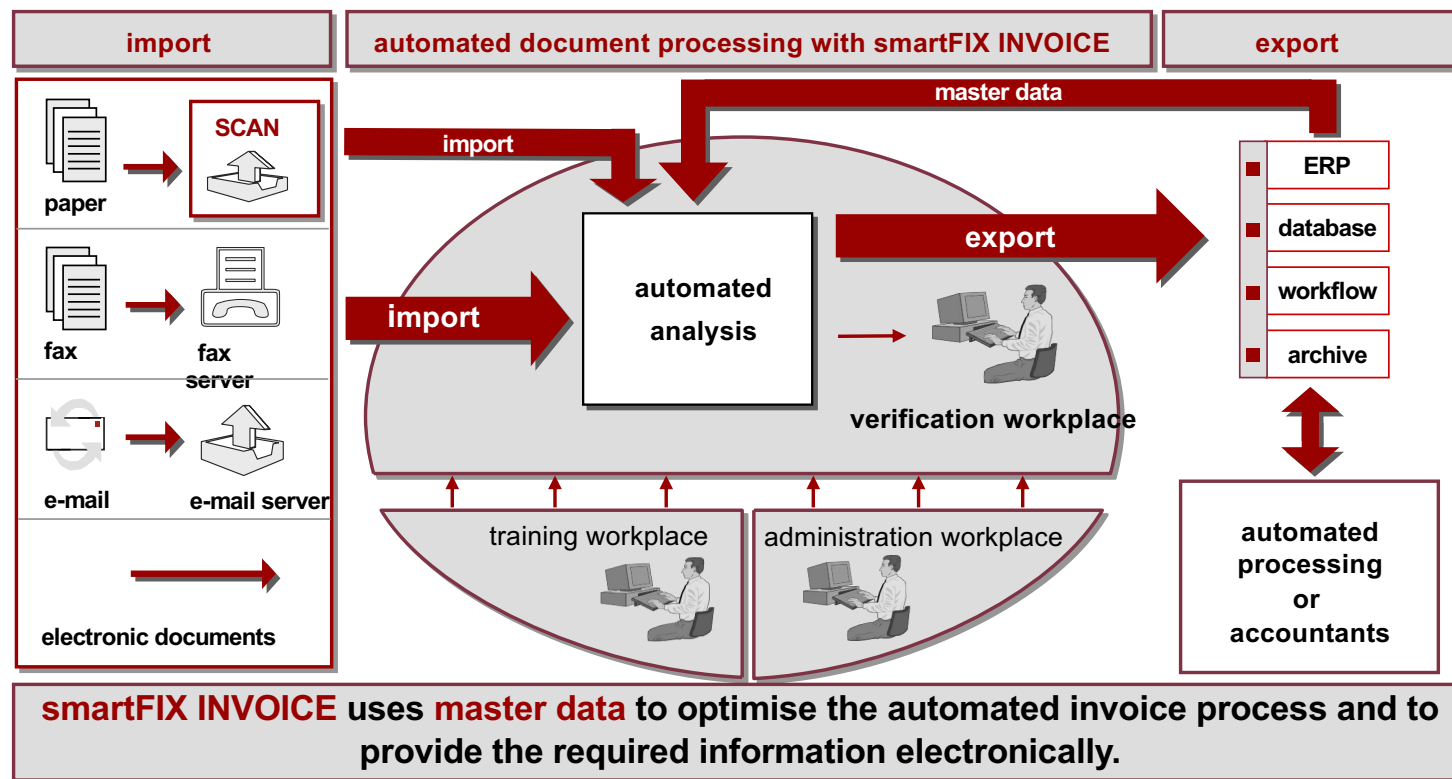
### DAS 2002

### Princeton



## The focus of our submissions after founding and running a company was much more on real systems

A. Dengel and B. Klein, *smartFIX: A Requirements-Driven System for Document Analysis and Understanding*, Proceedings DAS2002., Princeton, NJ, USA (Aug. 2002), pp. 433-444.



## At DAS 2004 we could feel the spirit of the medieval palaces in Florence

*Simone Marinai, Andreas Dengel:*

Document Analysis Systems VI, 6th International Workshop, DAS 2004, Florence, Italy, September 8-10, 2004, Proceedings. Lecture Notes in Computer Science 3163, Springer 2004, ISBN 3-540-23060-2

# DAS 2004

# Florence



B. Klein, S. Agne, and A. Dengel, *Results of a Study of Invoice Reading Systems in Germany*, Proceedings DAS2004, Florence, Italy, (Sept. 2004), pp. 451-462.

# The conference trip brought us to New Zealand for DAS 2006



*Horst Bunke, A. Lawrence Spitz:*

Document Analysis Systems VII, 7th International Workshop, DAS 2006, Nelson, New Zealand, February 13-15, 2006, Proceedings. Lecture Notes in Computer Science 3872, Springer 2006, ISBN 3-540-32140-3

## DAS 2006

## Nelson



B. Klein, S. Agne, and A. Dengel, *Benchmarking of invoice analysis systems*, Proceedings DAS2006, Nelson, NZ (Feb. 2006), pp. 312-323. .

# DAS 2008 was hosted at the world heritage of Nara, Japan



*Koichi Kise, Hiroshi Sako:*

The Eighth IAPR International Workshop on Document Analysis Systems, DAS 2008, September 16-19, 2008, Nara, Japan. IEEE Computer Society 2008, ISBN 978-0-7695-3337-7

## DAS 2008

## Nara



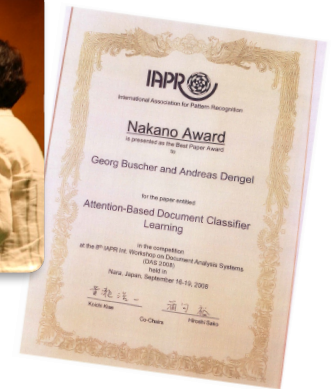
Image Source: <https://www.urlaubstracker.de/nara-tipps/>

©andreas.dengel@dfki.de - 2022 - p. 23

# By combining gaze paths and document analysis in our DAS 2008 paper, we could significantly improve classification and retrieval

G. Buscher and A. Dengel, *Attention-Based Document Classifier Learning*, Proceedings DAS2008, Nara, Japan, (Sep. 2008), pp. 87-94

**Text 2.0** (<https://www.youtube.com/watch?v=8QocWsWd7fc>)



We identified saccadic gaze feature to distinguish **reading** and **skimming** for individual users to **define textual relevance**



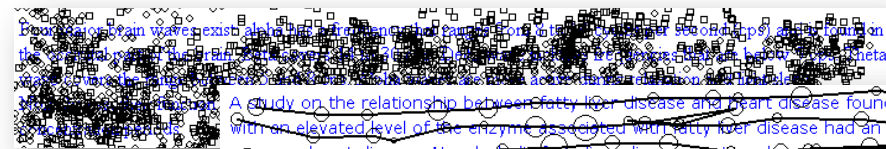
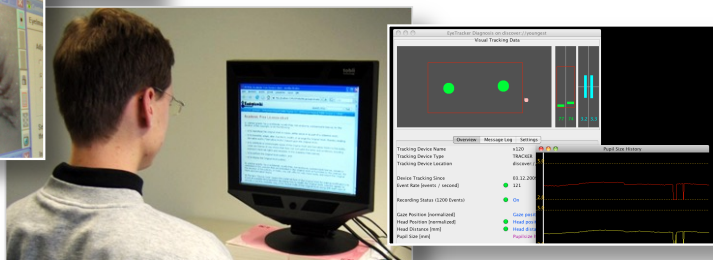
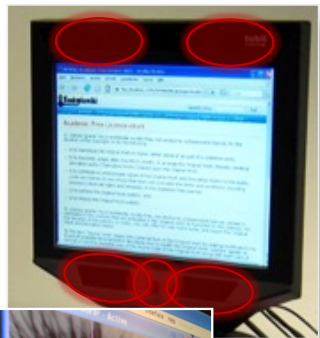
For building better classifiers, we tried to understand how users classify documents based on their reading behavior on specific text



We analyzed how users read search results to implement implicit relevance feedback to the search tool

**45% better!**

**20% better!**



A study on the relationship between fatty liver disease and heart disease found that patients with an elevated level of the enzyme associated with fatty liver disease had an increased risk of coronary heart disease. Nonalcoholic fatty liver disease is strongly associated with predictors of heart disease, such as high cholesterol, insulin resistance, obesity and central fat distribution. Patients with this disease are therefore expected to have an increased risk of heart disease.

## Our American friends provided a productive atmosphere at DAS 2010

*David S. Doermann, Venu Govindaraju, Daniel P. Lopresti, Premkumar Natarajan:*

The Ninth IAPR International Workshop on Document Analysis Systems, DAS 2010, June 9-11, 2010, Boston, Massachusetts, USA. ACM International Conference Proceeding Series, ACM 2010, ISBN 978-1-60558-773-8

# DAS 2010

## Boston



By combining gaze paths and document analysis, we could significantly improve classification and retrieval

M. Liwicki, H.M. S. M. El-Neklawy, and A. Dengel, *Touch & Write – A Multitouch Table with Pen-Input*, Proceedings DAS10, Boston, MA, USA, (June 2010), pp. 479-484.



## DAS 2012 was an ideal place to combine work and fun



*Michael Blumenstein, Umapada Pal, Seiichi Uchida:*

10th IAPR International Workshop on Document Analysis Systems, DAS 2012, Gold Coast, Queensland, Australia, March 27-29, 2012. IEEE Computer Society 2012, ISBN 978-1-4673-0868-7

# DAS 2012

# Gold Coast



## By combining gaze paths and document analysis, we could significantly improve classification and retrieval

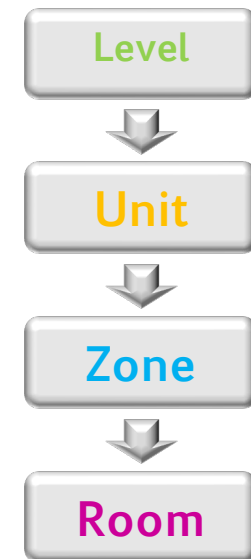
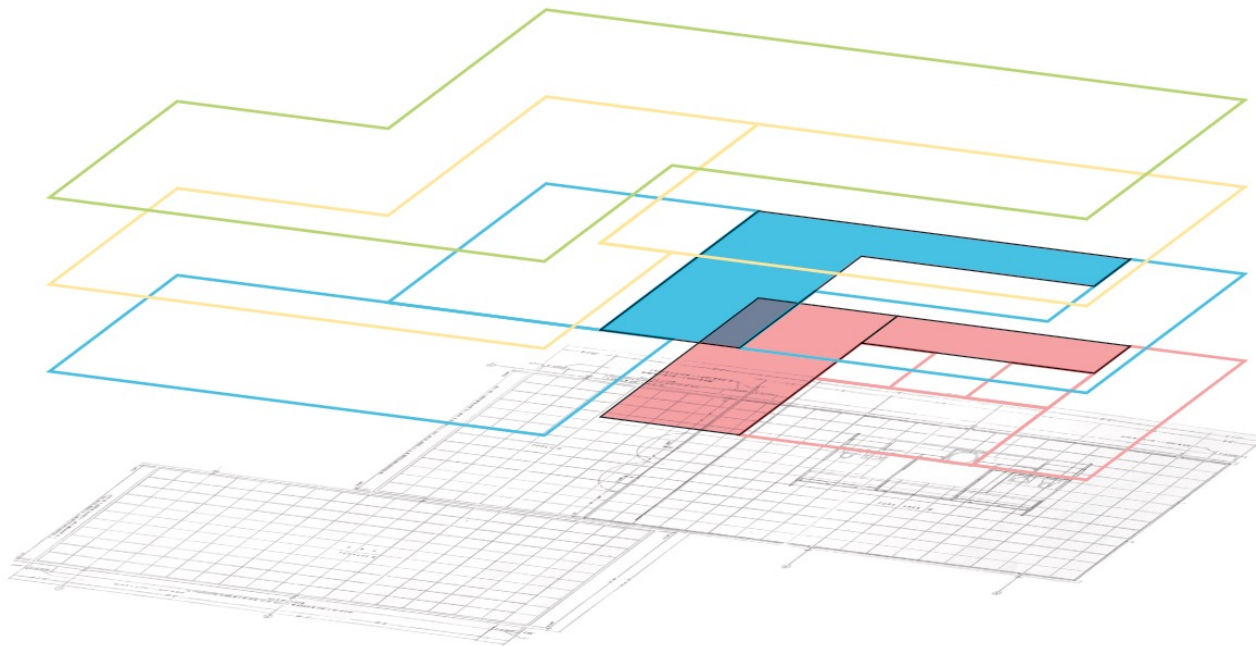
S. Ahmed, M. Liwicki, M. Weber, and A. Dengel, *Automatic Room Detection and Room Labeling from Architectural Floor Plans*,  
Proceedings DAS2012, Gold Coast, Australia (March 2012), pp. 338-342.

### *a.SCatch*

A typical floor plan is composed of walls, pictorial icons and designators describing room functions






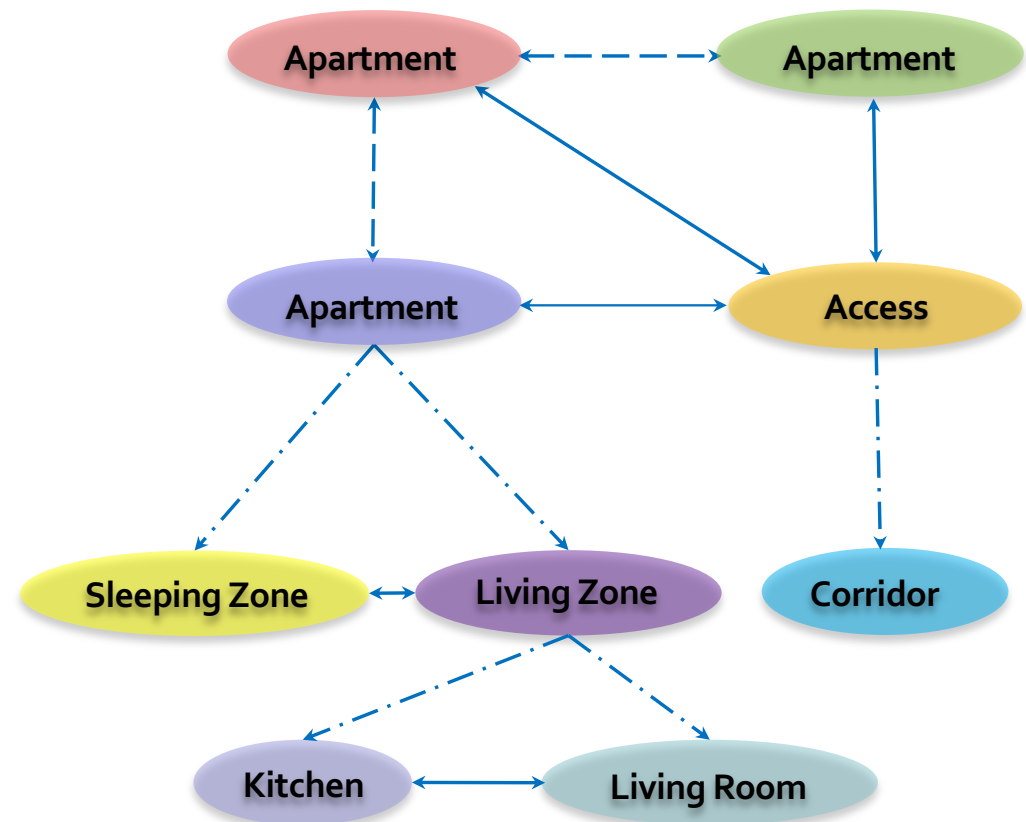
In *a.SCatch* we employed a semantic representation of floor plans based on the aggregation of living spaces



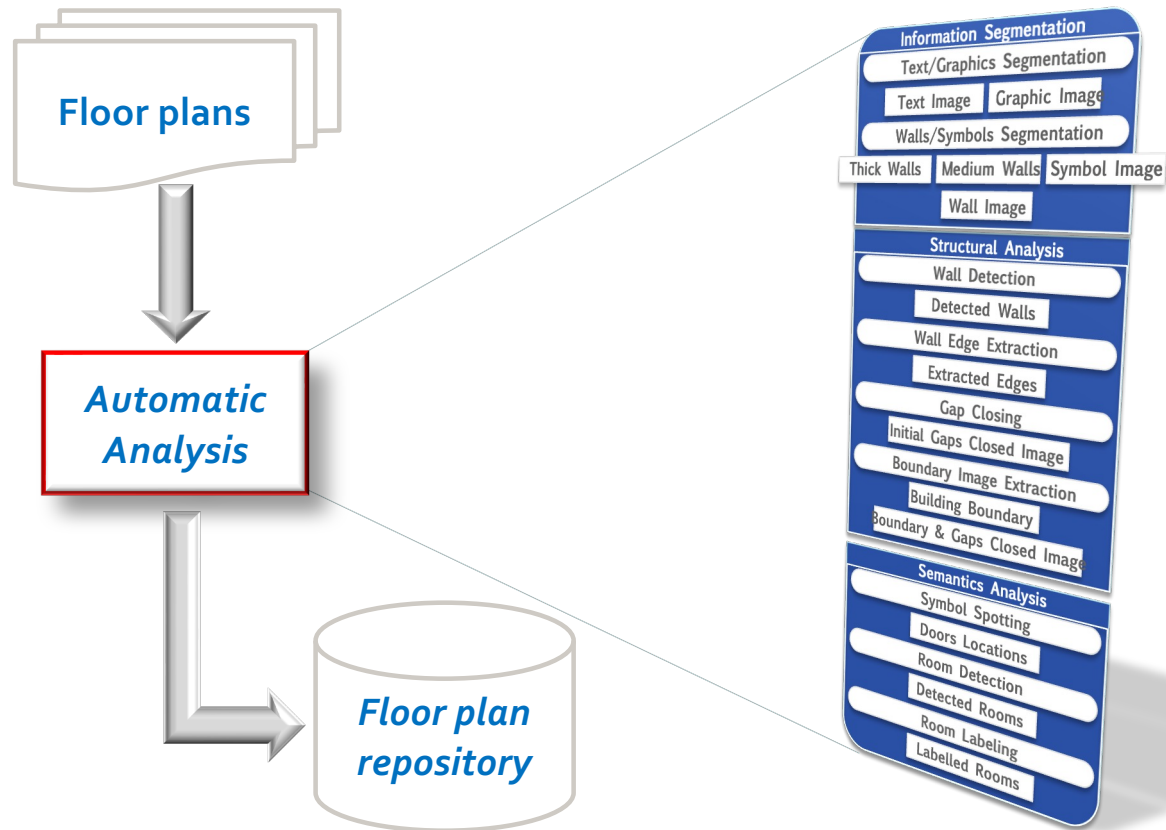
We may further describe the spatial relationships among the entities of a living space by semantic links



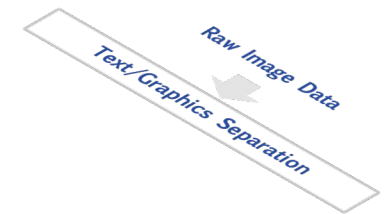
-  has part
-  direct connection
-  adjacent connection



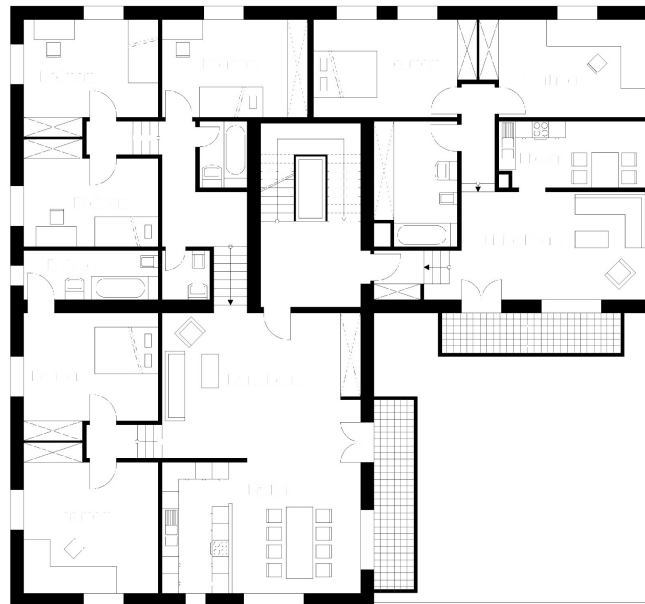
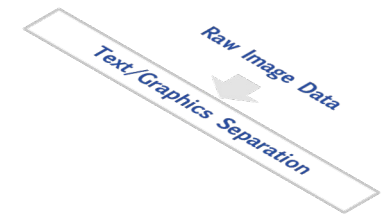
In a first step, floor plans have to be scanned, analyzed, and transformed into an semantic representation



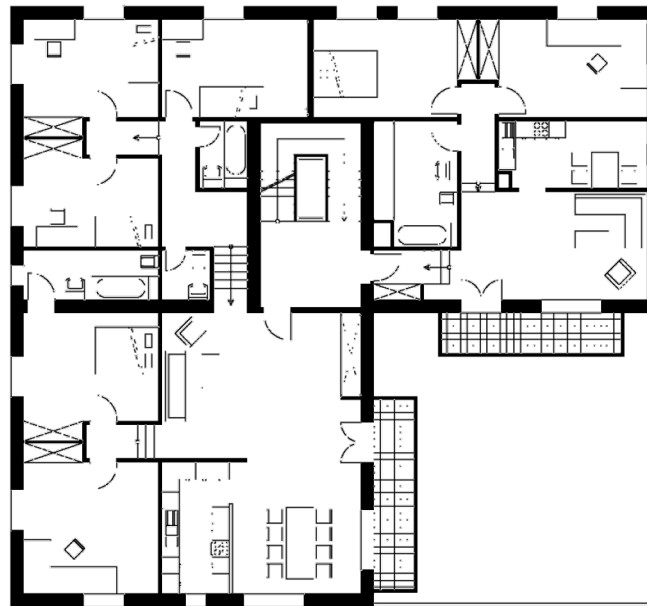
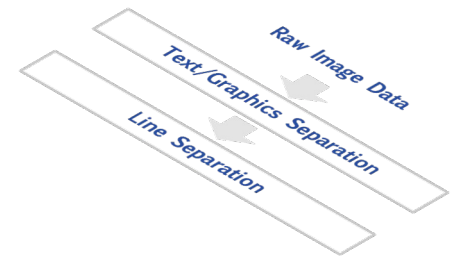
In order to select appropriate recognition techniques an important step is an initial text/graphics separation



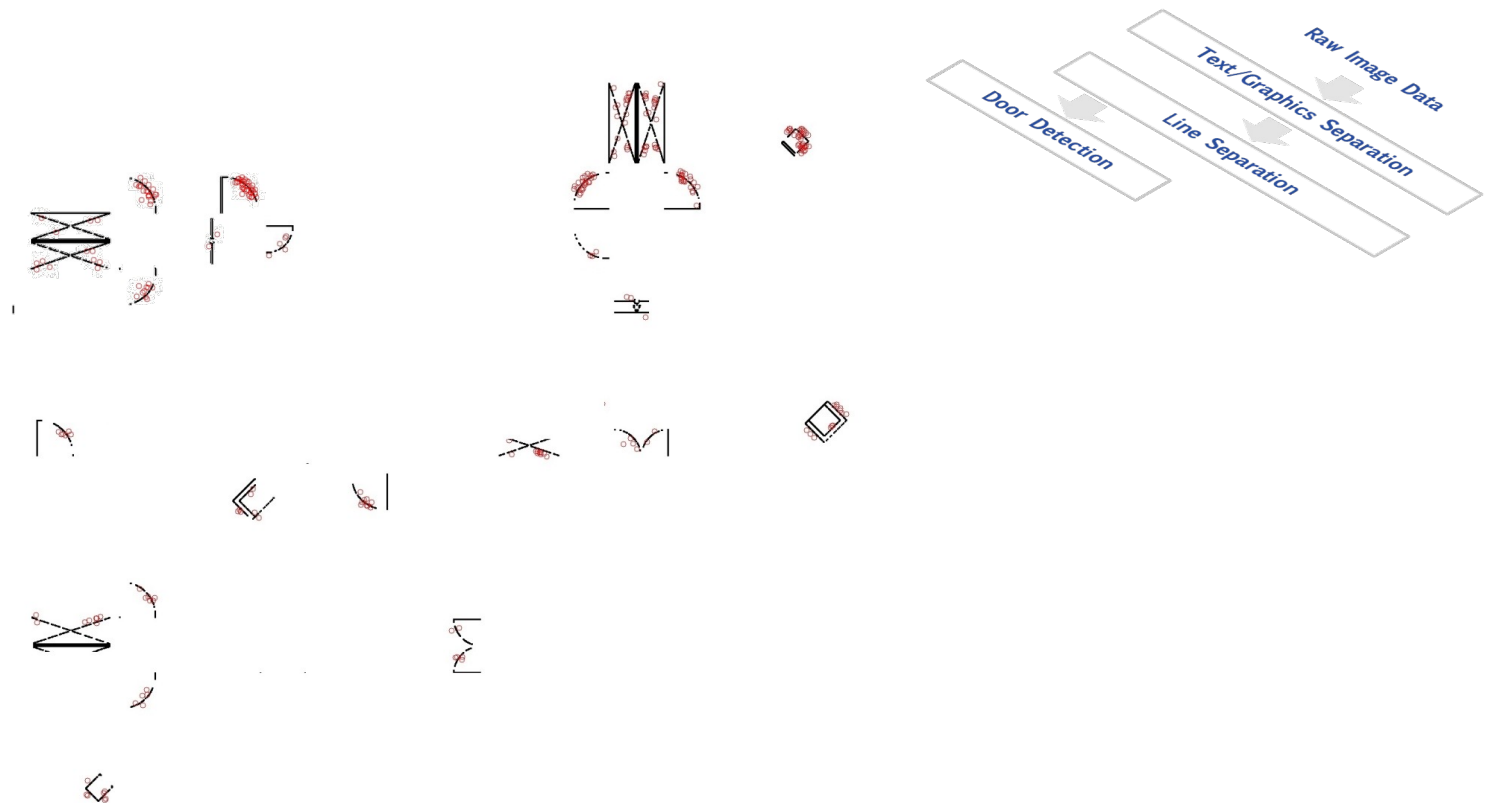
In order to select appropriate recognition techniques an important step is an initial text/graphics separation



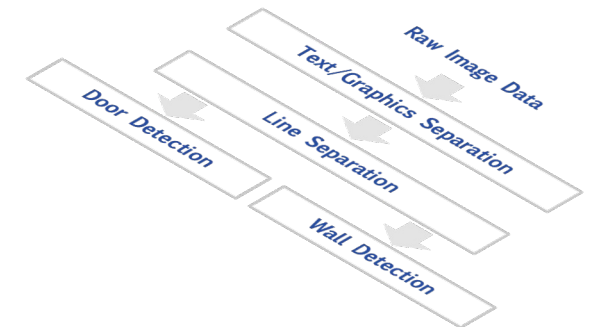
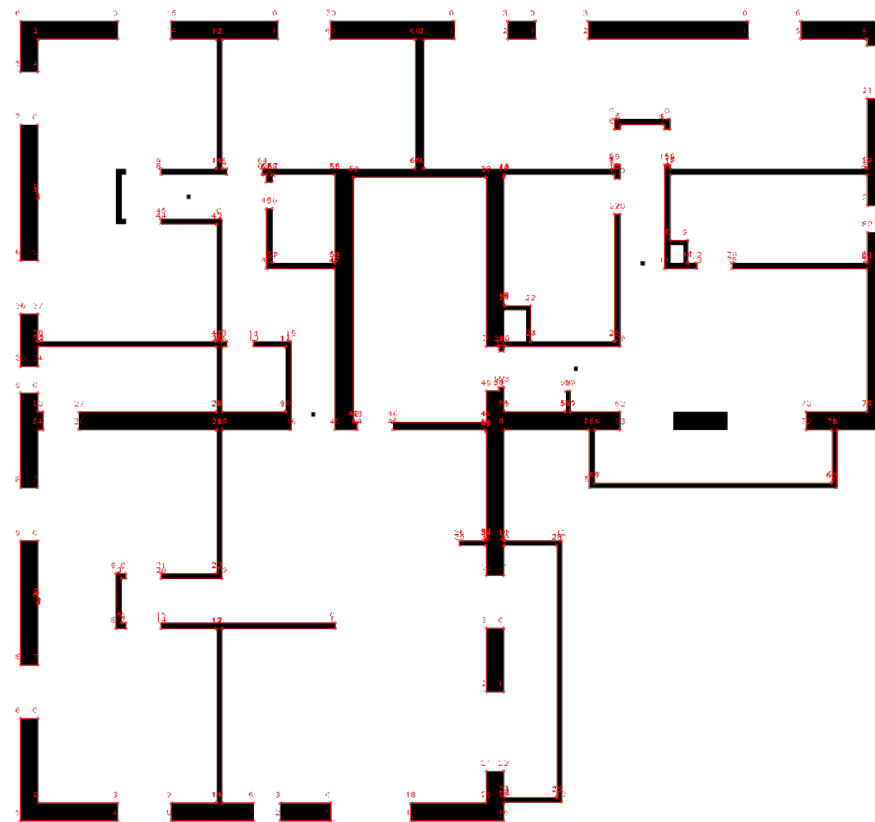
In the resulting drawing, the various types of lines are separated



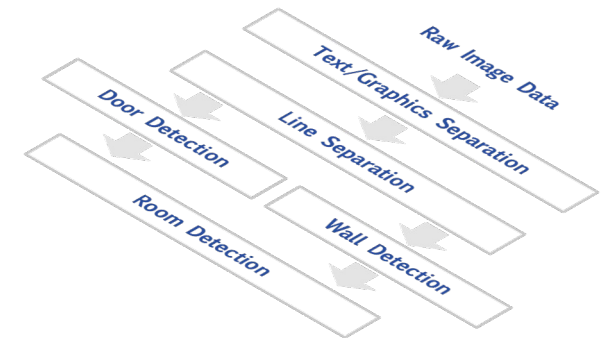
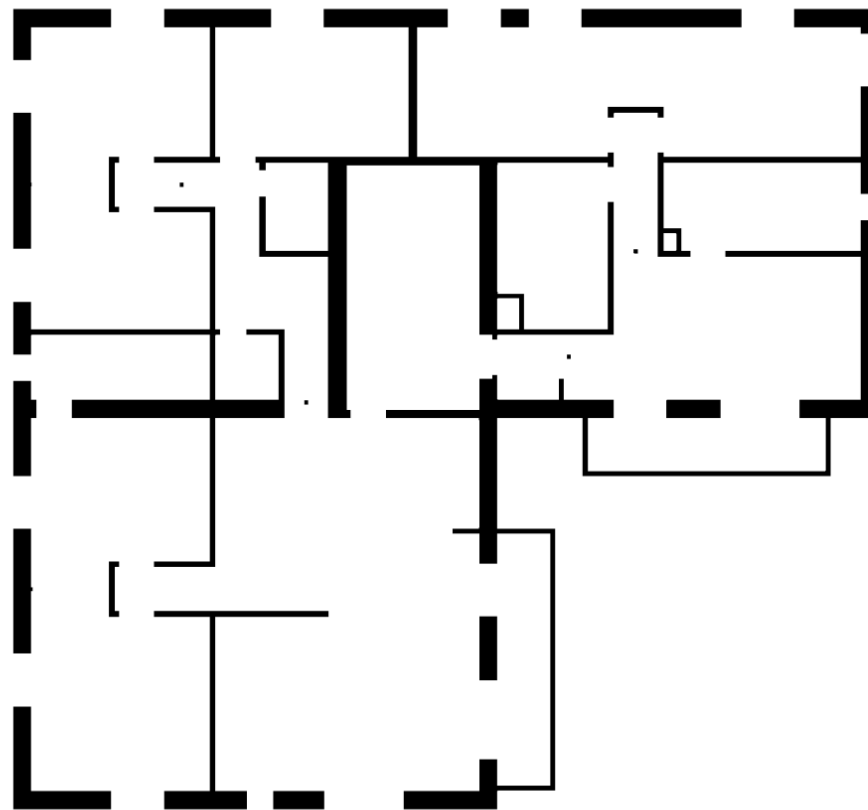
In the next step, *a.SCatch* attempts to detect the doors captured in the drawing



Furthermore, the walls have to be detected



Finally the rooms have to be recognized as basic living spaces



↔ direct connection  
- - - adjacent connection

# DAS 2016 was hosted at one of the most spectacular places I visited



*Apostolos Antonacopoulos, Basilis Gatos:*

12th IAPR Workshop on Document Analysis Systems, DAS 2016, Santorini, Greece, April 11-14, 2016. IEEE Computer Society 2016, ISBN 978-1-5090-1792-8

## DAS 2016

## Santorini



# At DAS 2016, we proposed a system for fake news recognition



S. Elkasrawi, A. Abdelsamad, S. Bukhari, M. Al-Naser, and A. Dengel, *What you see is what you get?*  
*Automatic Image Verification for Online News Contents*,  
Proceedings DAS 2016, Santorin, Greece (Apr. 2016), pp. 114-119

April 17th, 2013

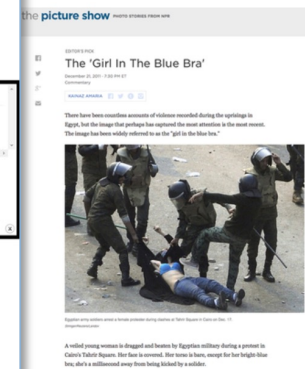
Tweet on Twitter shows alleged assaults by Venezuelan National Guard during peaceful protests



Revealed by the  
DFKI NewsVerifier!

December 21st, 2011

npr.org publishes article about spring revolution in Egypt. Original photo shows woman being kicked and pulled away by Egyptian soldiers



**NewsVerifier**

Image fraud detection in social media



At DAS 2014, we could enjoy the warm spring in ancient Loire valley



*Jean-Yves Ramel, Marcus Liwicki, Jean-Marc Ogier, Koichi Kise, Ray Smith:*

11th IAPR International Workshop on Document Analysis Systems, DAS 2014, Tours, France, April 7-10, 2014.


IEEE Computer Society 2014, ISBN 978-1-4799-3244-3

**DAS 2014**

**Tours**



## Recognition



This comparison shows two printed versions of the letter 'A'. The left version, labeled 'ink jet printing', is blurry and has a soft, fuzzy appearance. The right version, labeled 'Laser printing', is sharp and has a clear, crisp appearance.

## Laser printing

## Analysis of CPS-Codes („Yellow Dots“)

## Statistical Layout Models

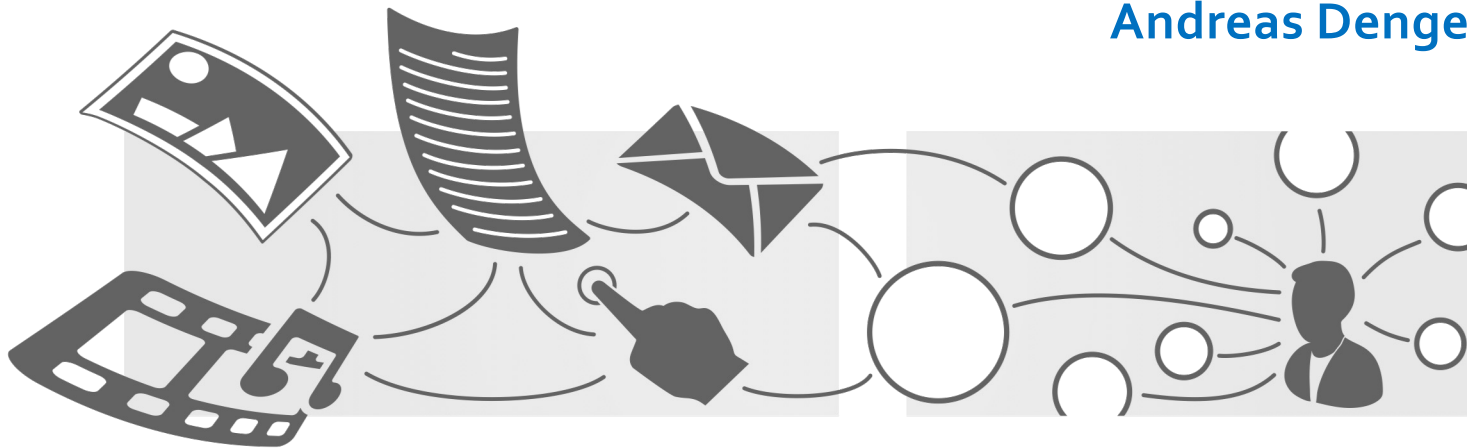
... and I gave a keynote on ...



# Document Evolution drives Document Analysis

– An Attempt for Stimulating a Discussion –

Andreas Dengel



DAS 2014, Tours, France

# In 2018, we proceeded with DAS in the wonderful Vienna



*Robert Sablatnig, Florian Kleber:*

13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018.

IEEE Computer Society 2018, ISBN 978-1-5386-3346-5

## DAS 2018

## Vienna

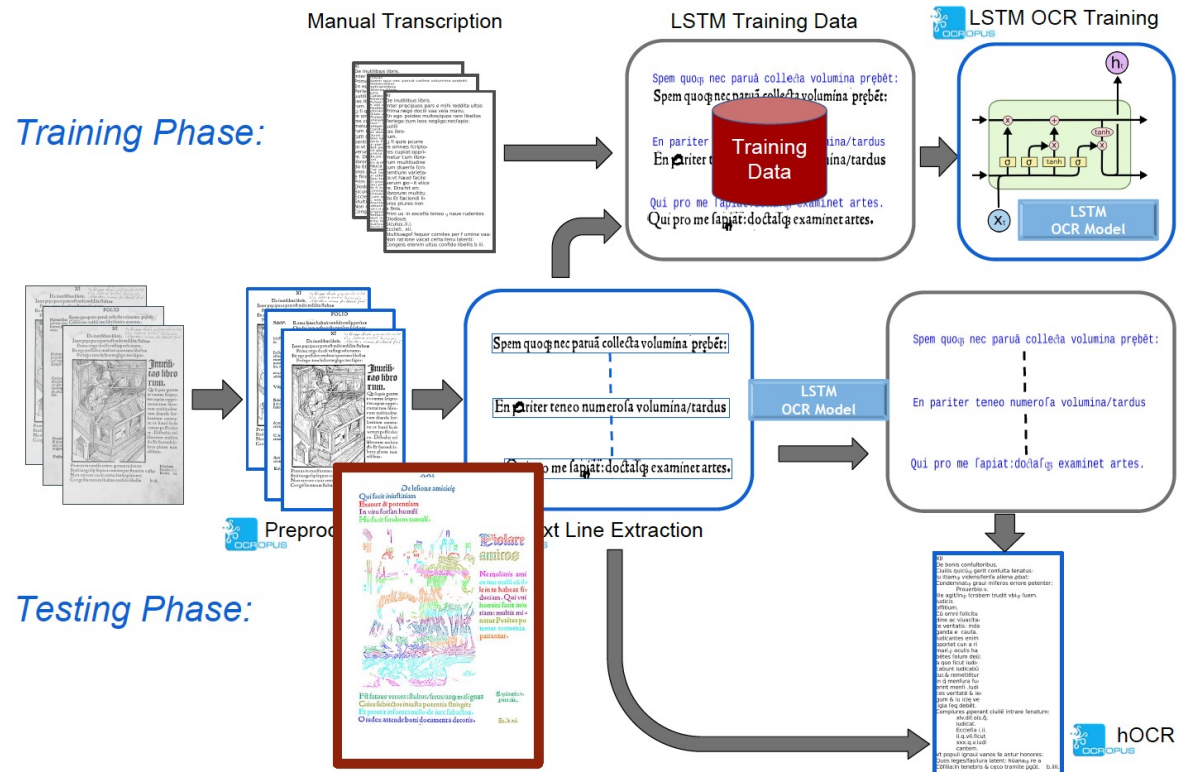


Image Source: <https://www.welt.de/reise/staedtereisen/article204092006/Liebeserklaerung-Warum-Wien-die-charmanteste-Stadt-ist.html>

©andreas.dengel@dfki.de - 2022 - p. 43

# Our text alignment method presented at DAS 2018 was part of a complex system analyzing historical documents

S. S. Bukhari, M. Saha, P.K. Giridhara, M.K. Lohano, and A. Dengel, anyalign: *An Intelligent and Interactive Text-Alignment Web-Application for Historical Documents*, Proceedings DAS 2018, Vienna, Austria (Apr. 2018).



**Kalimachos**

anyOCR and anyalign as part of library research and information system

In 2020, DAS was hosted by our colleagues in Wuhan, China



*Xiang Bai, Dimosthenis Karatzas, Daniel Lopresti:*

Document Analysis Systems - 14th IAPR International Workshop, DAS 2020, Wuhan, China, July 26-29, 2020, Proceedings. Lecture Notes in Computer Science 12116, Springer 2020, ISBN 978-3-030-57057-6

**DAS 2020**

**Wuhan**



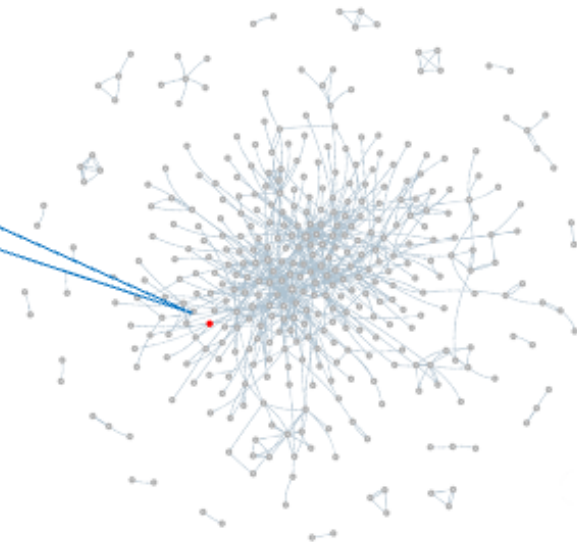
## By considering authority mining in conference proceedings, we expanded our view to document analysis systems at DAS 2020

M. Beck, S.T. Rizvi, A. Dengel, and S. Ahmed, *From Automatic Keyword Detection to Ontology-based Topic Modeling*, Proceedings DAS 2020, 14<sup>th</sup> Int'l IAPR Workshop on Document Analysis Systems, Wuhan, China (May 2020).



- ⇒ The goal is to reveal
- ... historical contributors, their relationship, topics, and technologies addressed by community members,
  - ... scientific trends and opinion leaders,
  - ... citation behavior as well as collaborating cliques

You!



The resulting Academic Community Explorer (ACE) is ready to be explored



3  
Communities 



4638  
Publications 



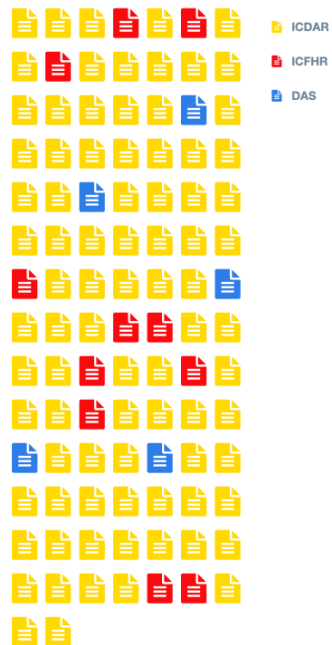
6255  
Authors 



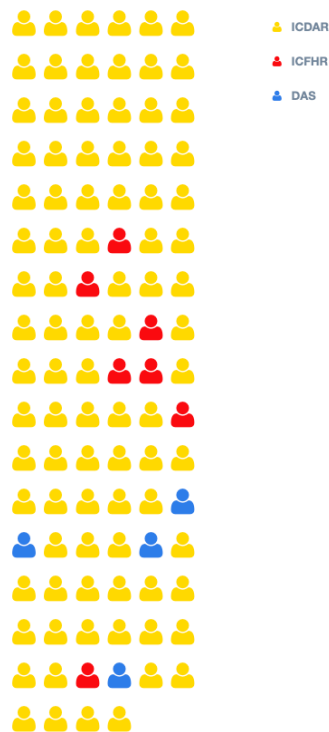
65775  
Citations 



Top 100 Most Cited Papers 



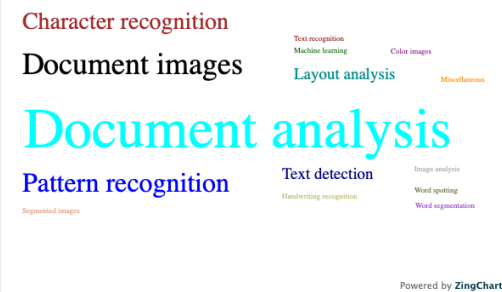
Top 100 Most Cited Authors 



Top 100 Most Influential Authors 



Top 15 Topics with Most Contributions 



Top 15 Topics with Most Citations 



<https://ace.opendfki.de>

# Thank you!



*Prof. Dr. Prof. h.c. Andreas Dengel  
DFKI GmbH  
Postfach 2080  
67608 Kaiserslautern  
email: [andreas.dengel@dfki.de](mailto:andreas.dengel@dfki.de)  
<http://www.dfki.de/~dengel>*

