UNIVERSITÄT BERN

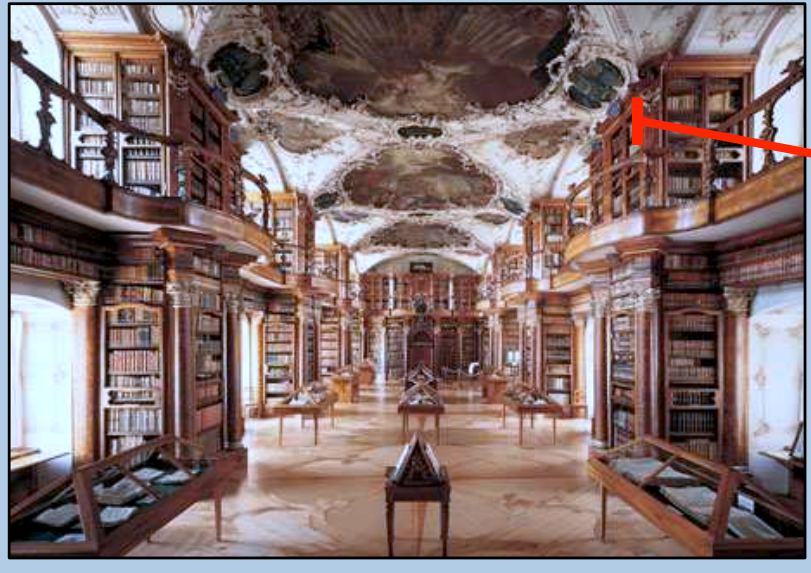
SEMINAR TALK LA ROCHELLE, JUNE 28

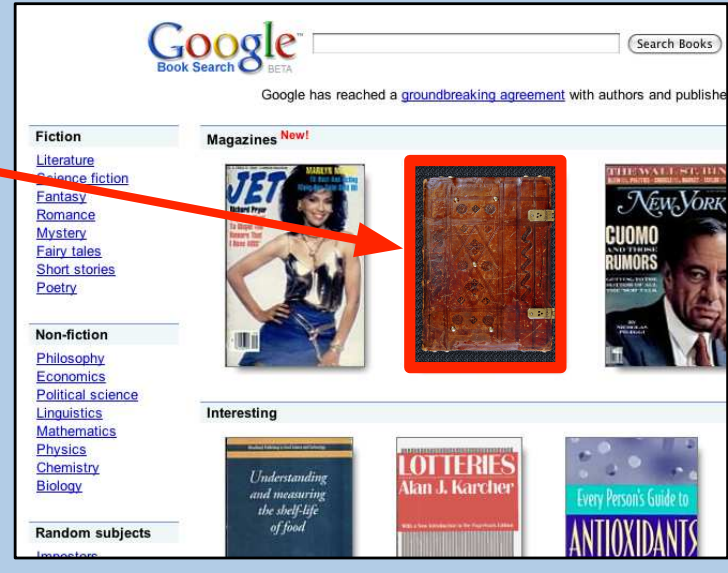
HANDWRITING RECOGNITION IN HISTORICAL DOCUMENTS

Andreas Fischer
Post-doctoral researcher, DIVA group
Department of Informatics, University of Fribourg
andreas.fischer@unifr.ch

Motivation

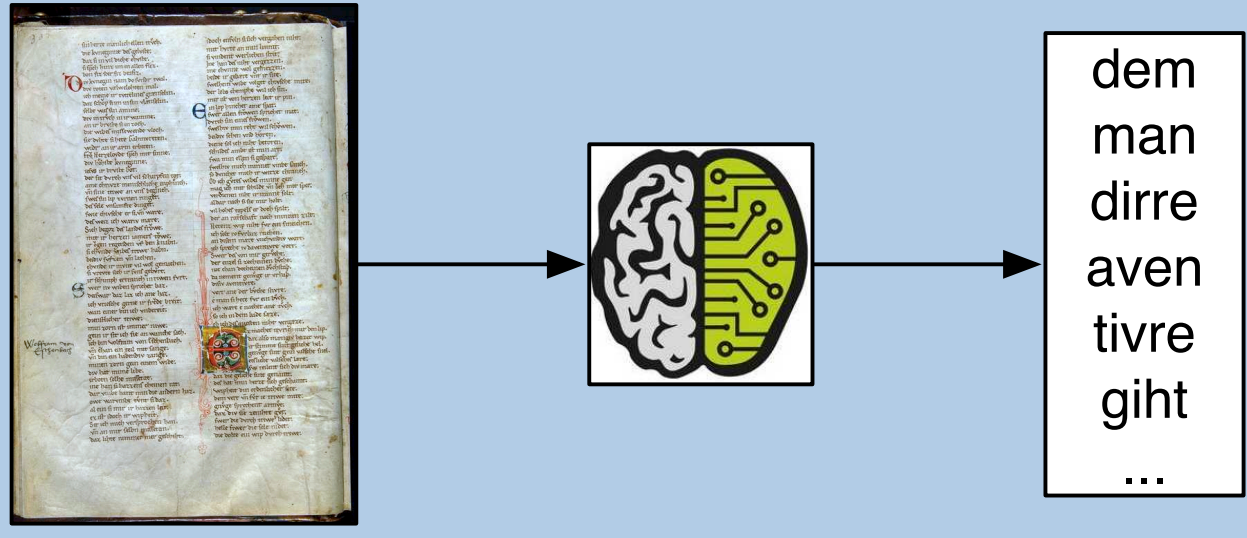
- Integration of historical manuscripts in digital libraries.
- Making our cultural heritage accessible to researchers and the public.





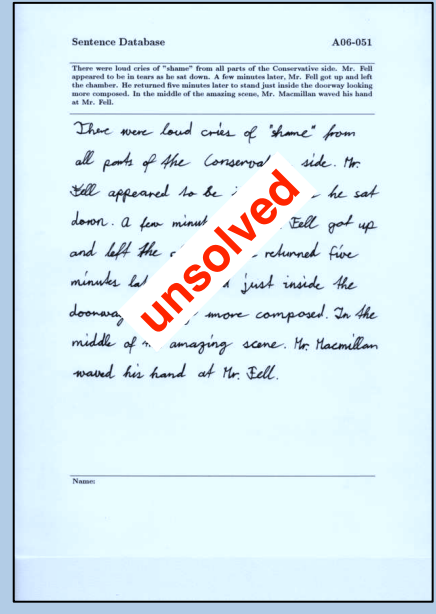
Objective

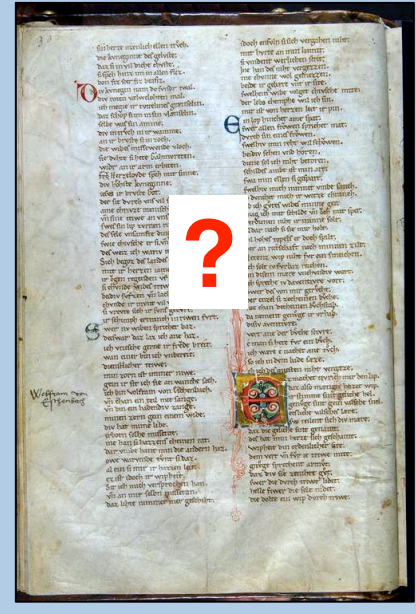
- Transcription of digital images into machine-readable text.
- Automatic reading needed for processing millions of manuscripts.



State of the Art

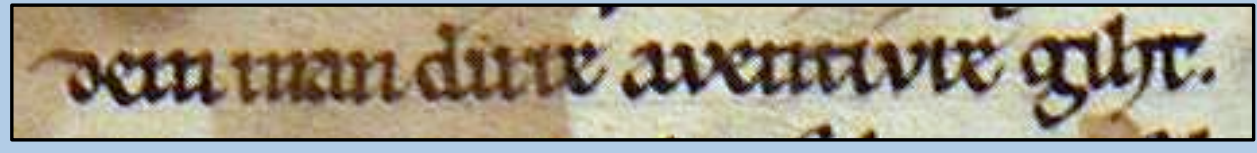
- Industrial solutions available for printed documents.
- Unsolved problem for handwriting images.





Learning by Samples

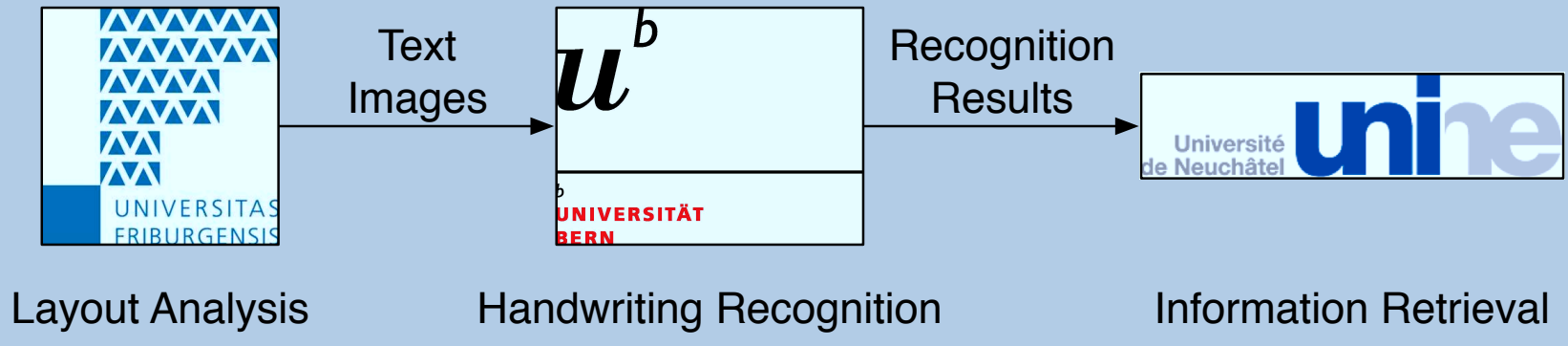
- Recognition rules are learned from samples.
- Difficult and costly to obtain in case of historical documents.



dem man dirre aventivre giht.

HisDoc Project

- Synergy research project of the Swiss National Science Foundation.
- Pioneering work on historical handwriting recognition.

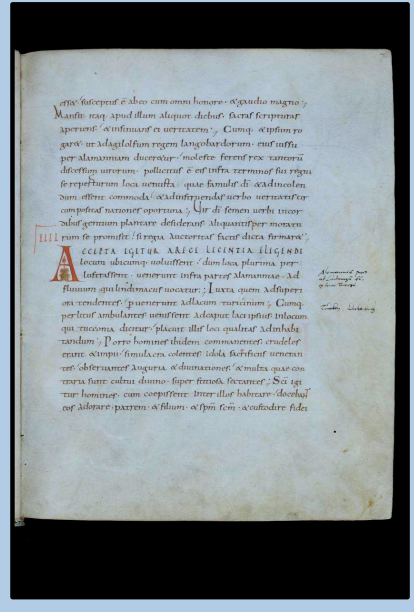


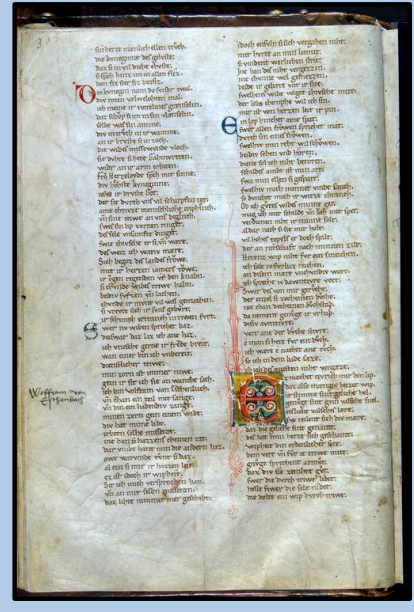
Outline

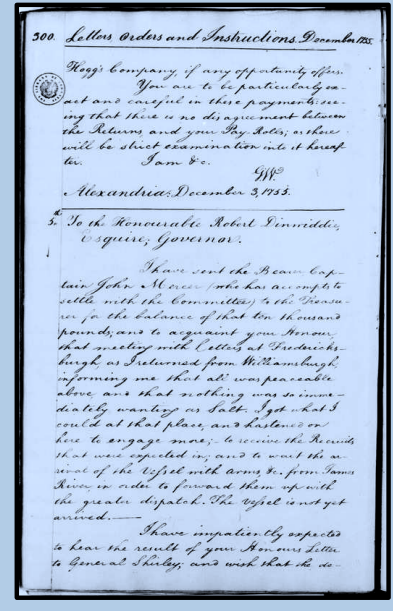
- **Database**
  - Ground Truth Creation
- Automatic Transcription
  - Graph Similarity Features
  - Fast Recognition Algorithm
- Keyword Spotting
- Transcription Alignment
- Conclusions

IAM-HistDB

- One of the first comprehensive research databases in the field.
- Freely available on the Internet.

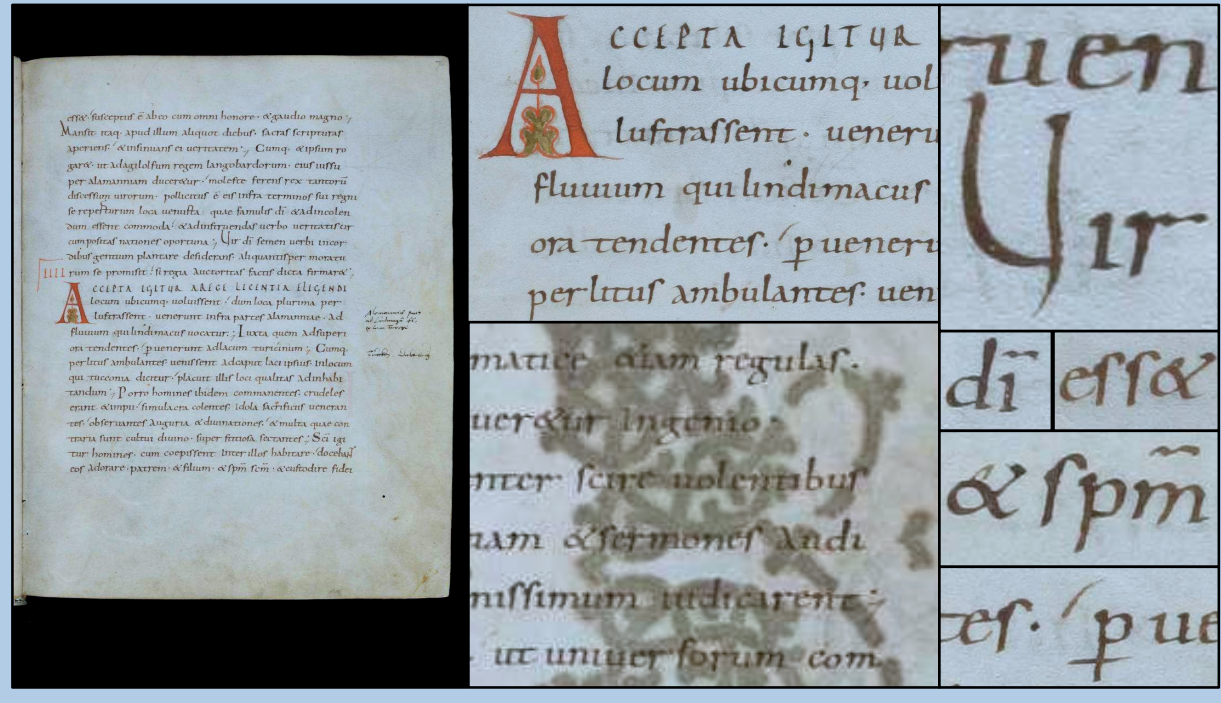






Saint Gall Database

- *Vita Sancti Galli* by Walafrid Strabo, Cod. Sang. 562, 9th century.
- 60 pages, Latin, Carolingian script.

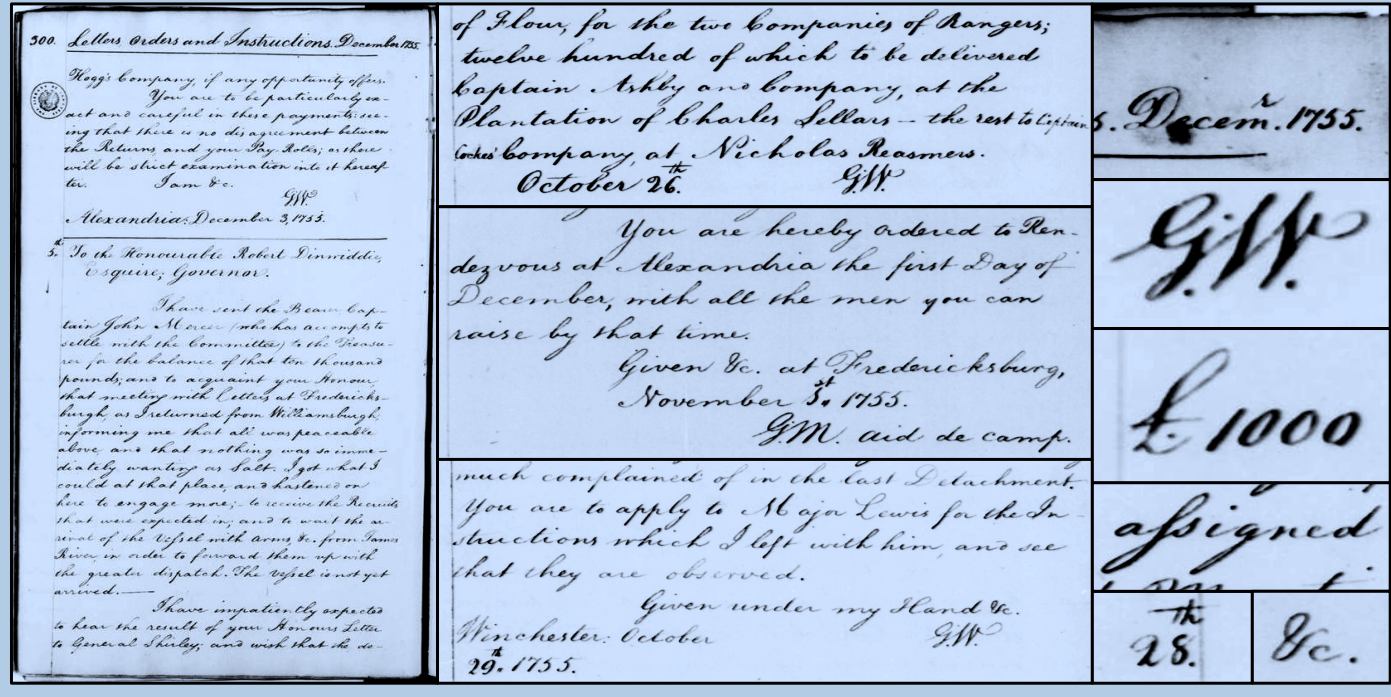


Parzival Database

- *Parzival* by Wolfram von Eschenbach, Cod. 857, 13th century.
- 47 pages, German, Gothic script.

George Washington Database

- Letters of George Washington, Library of Congress, 18th century.
- 20 pages, English, longhand script.

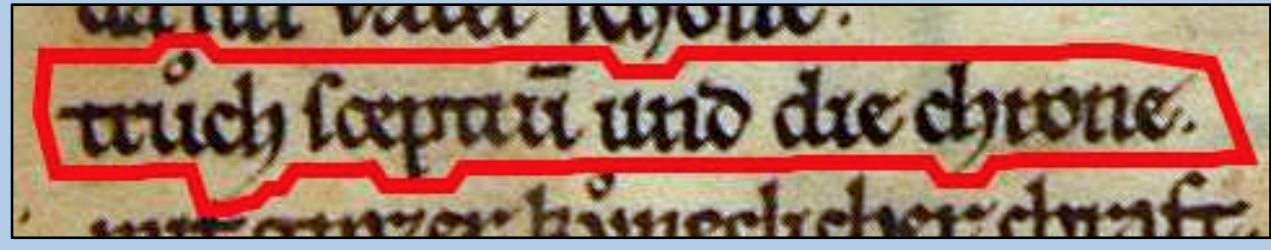


Outline

- Database
  - **Ground Truth Creation**
- Automatic Transcription
  - Graph Similarity Features
  - Fast Recognition Algorithm
- Keyword Spotting
- Transcription Alignment
- Conclusions

Ground Truth

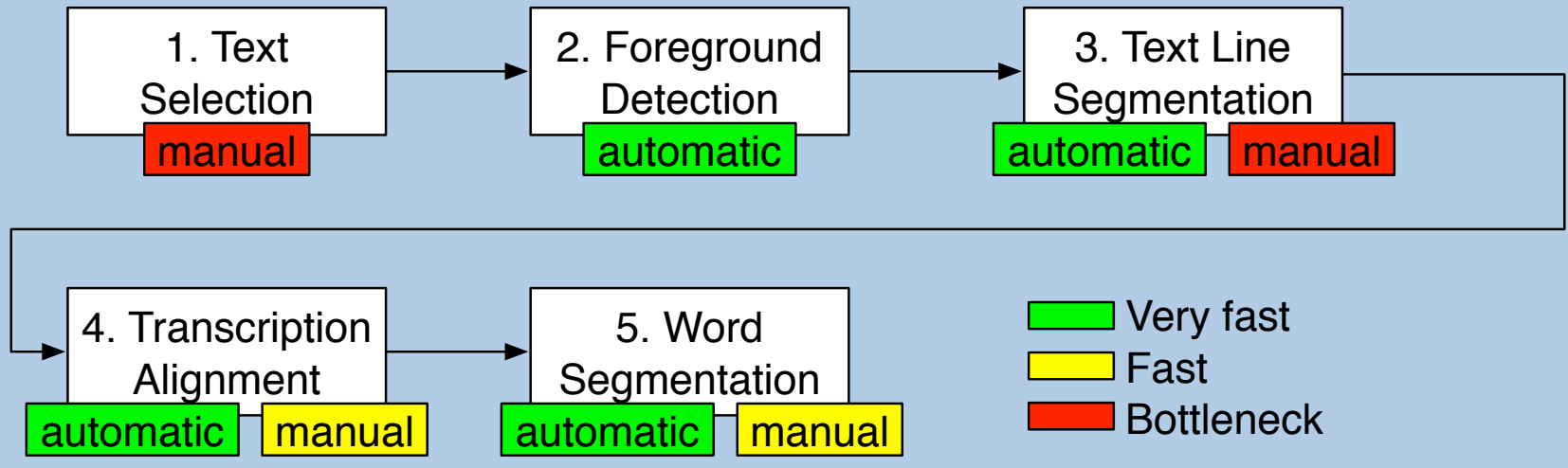
- Supervised manuscript annotation with correct transcriptions.
- Enables learning by samples as well as performance evaluation.



trůch sceptrū(sceptrum) und die chrone.

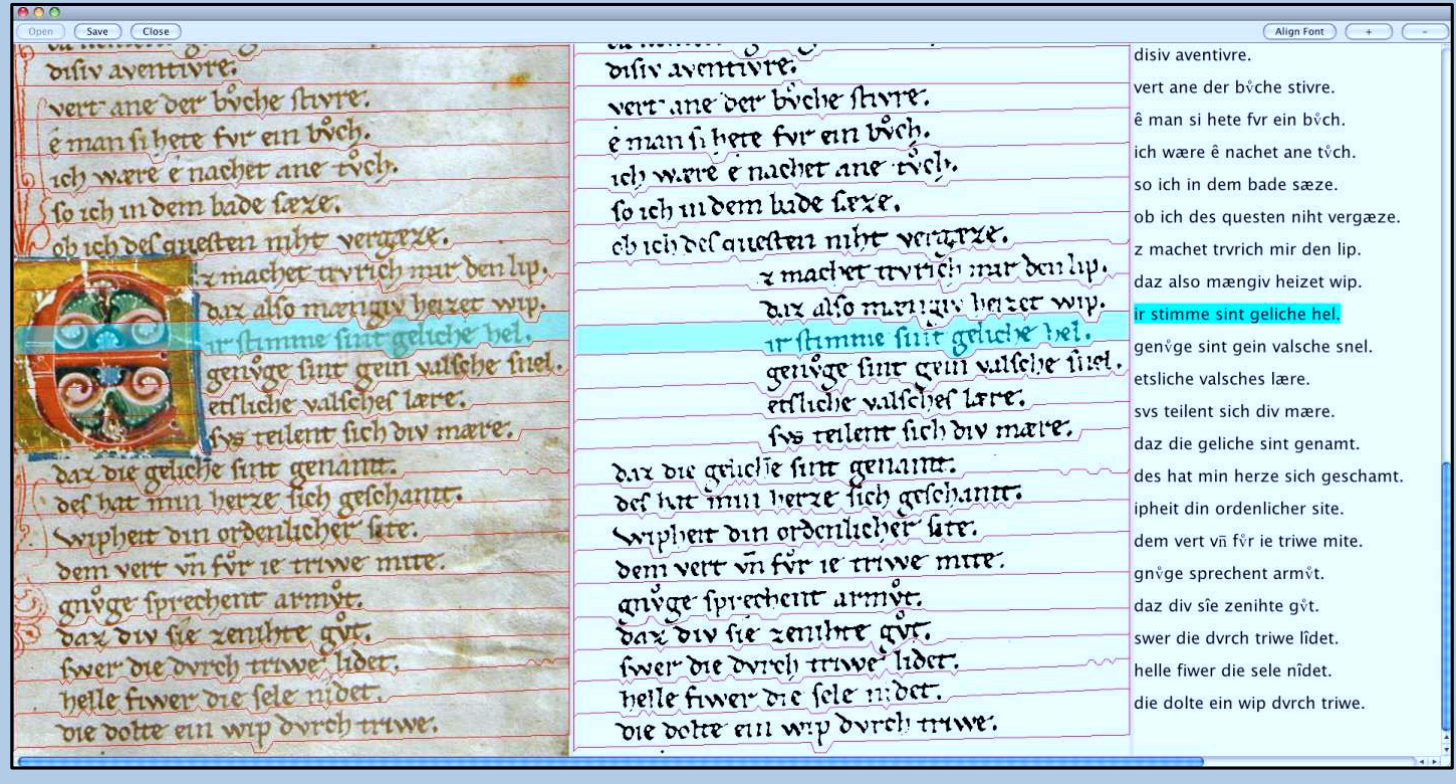
Semi-Automatic Proceeding

- Automatic: image segmentation (Indermühle *et al.*, 2009) & alignment.
- Manual: text area selection and corrections.



Human $\leftrightarrow$ Machine Interface

- Graphical user interface for interactive correction.
- Laypersons spent about 10 minutes for one page.

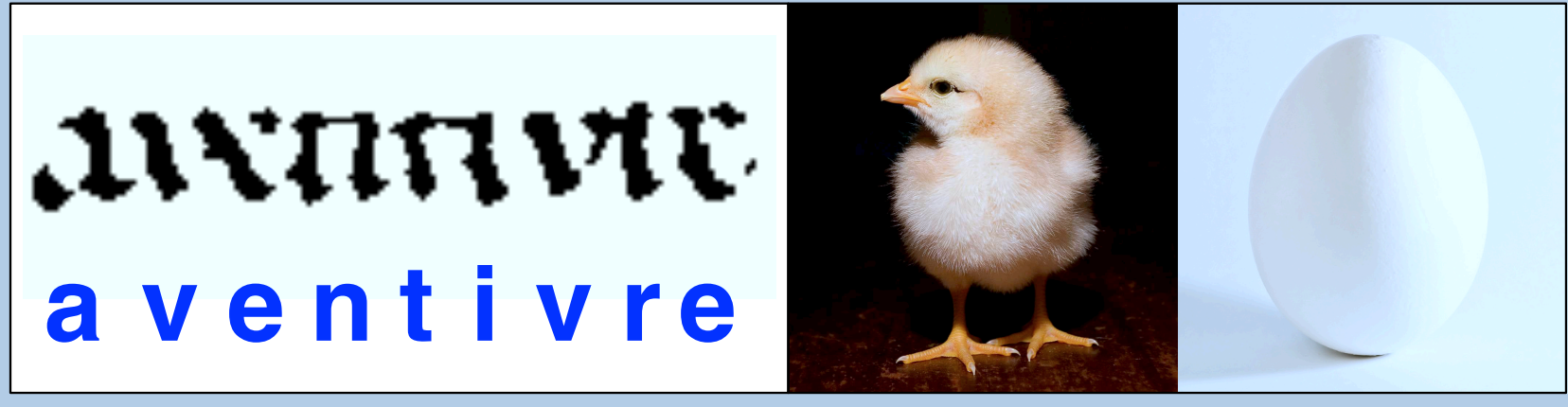


Outline

- Database
  - Ground Truth Creation
- **Automatic Transcription**
  - Graph Similarity Features
  - Fast Recognition Algorithm
- Keyword Spotting
- Transcription Alignment
- Conclusions

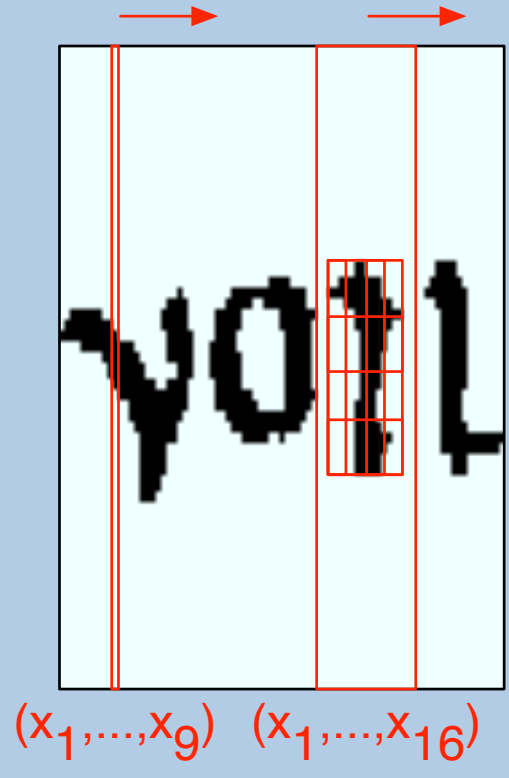
Sayre's Paradox (1973)

- Recognition requires character segmentation.
- Character segmentation requires recognition.



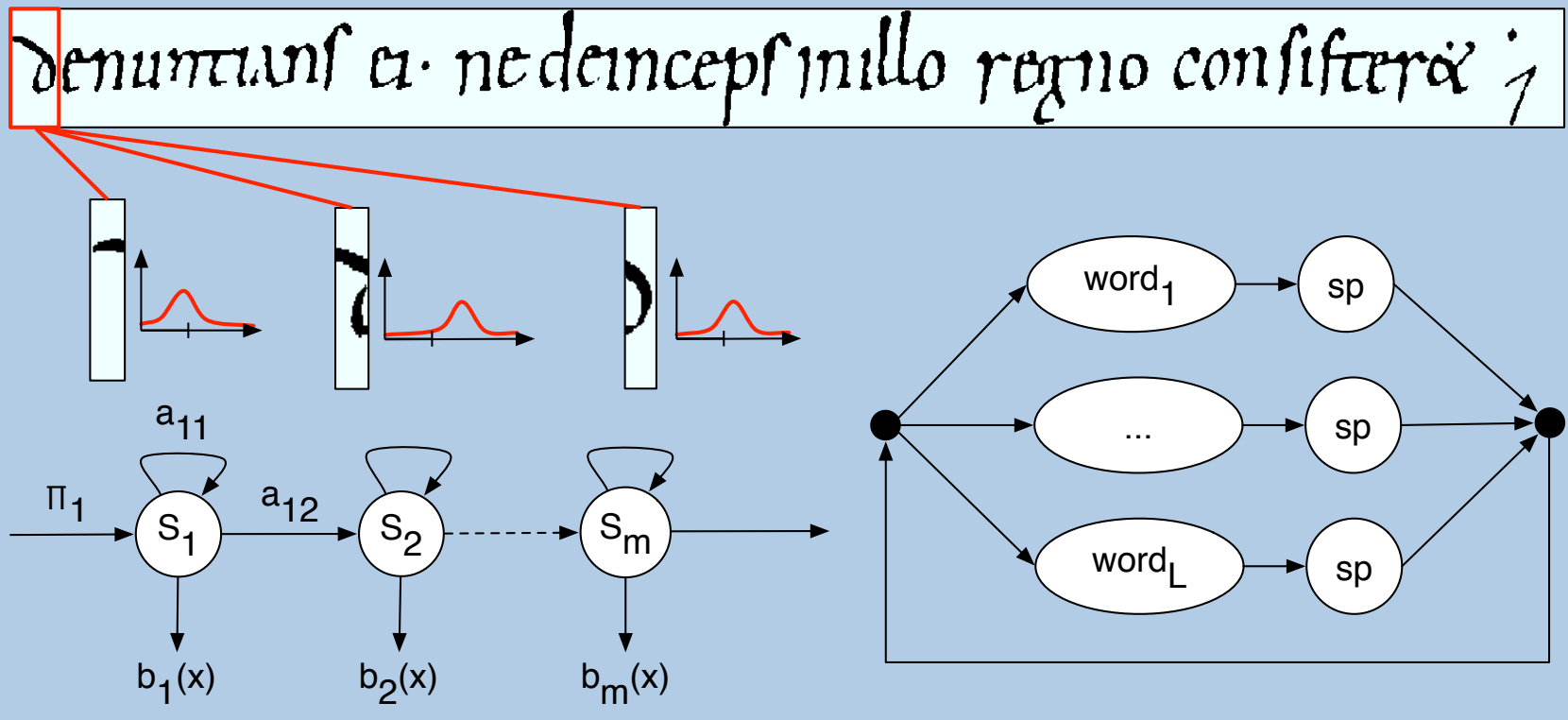
Sliding Window Approach

- Slice into sub-character parts, extract features (Marti & Bunke, 2001).
- Reconstruct characters during recognition.



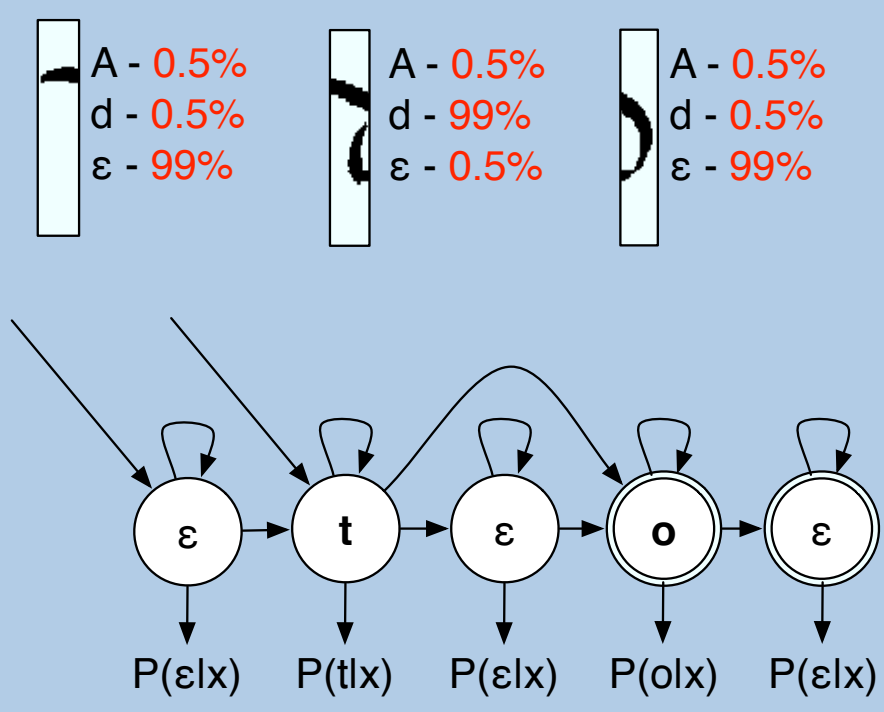
Hidden Markov Models (HMM)

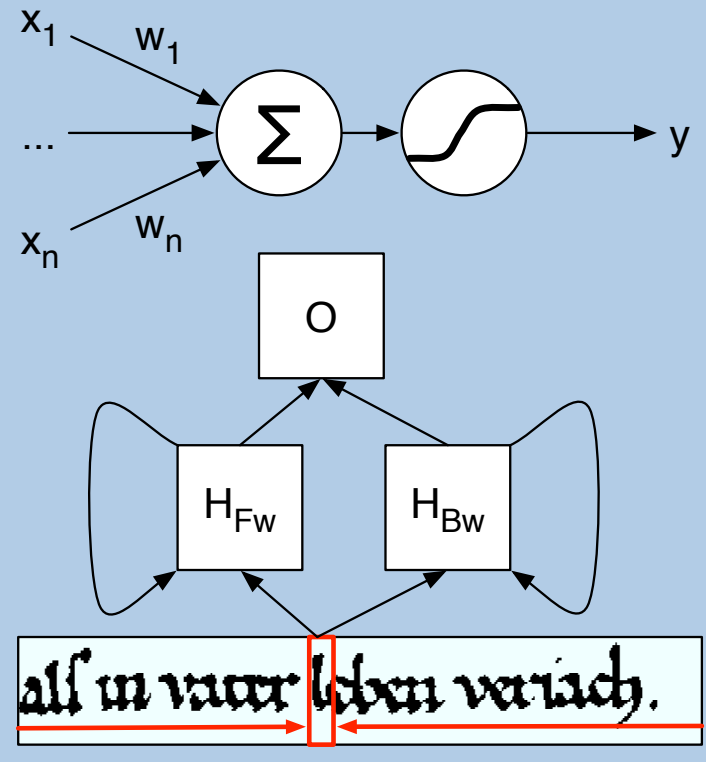
- Generative approach to model feature distributions (Rabiner, 1989).
- Efficient algorithms for training (Baum-Welch) and recognition (Viterbi).



Neural Networks (NN)

- Discriminative approach to distinguish characters (Graves *et al.*, 2009).
- Analysis reveals: training and recognition closely related to HMMs.

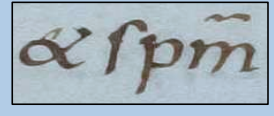




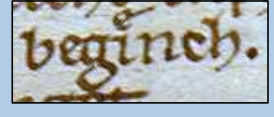
Text Line Recognition Results

- Single manuscript recognition, closed vocabulary, word bigrams.
- NN word error: 6% $SG_{0.5}$, 7% $PAR_{2.2}$, 18% $GW_{0.3}$

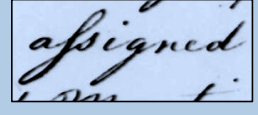
→ Significantly better than modern scripts, promising for digital libraries.



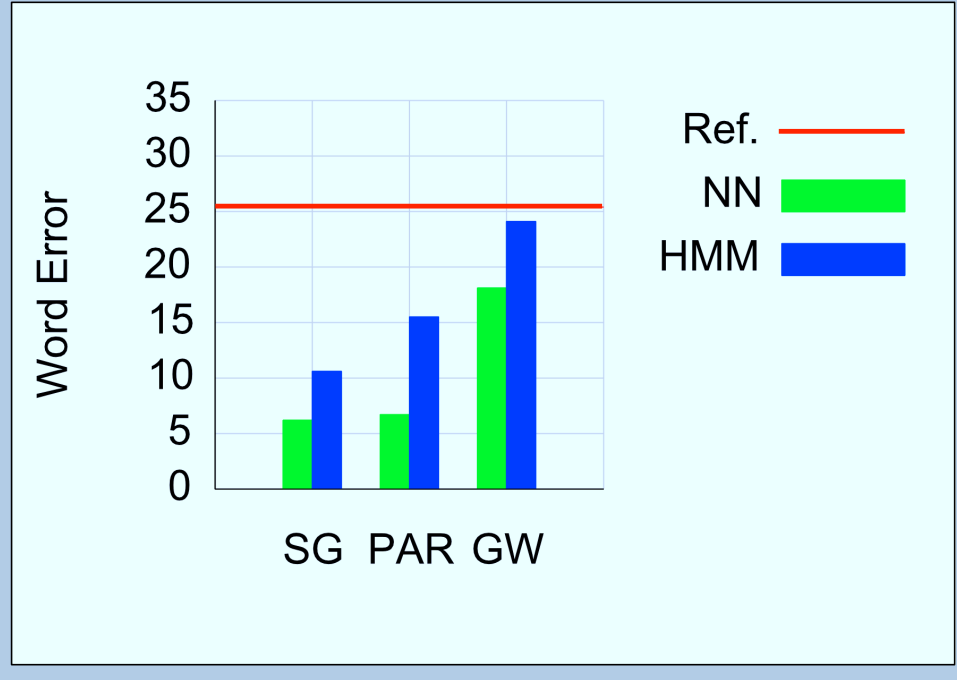
$SG_{0.5}$



$PAR_{2.2}$



$GW_{0.3}$

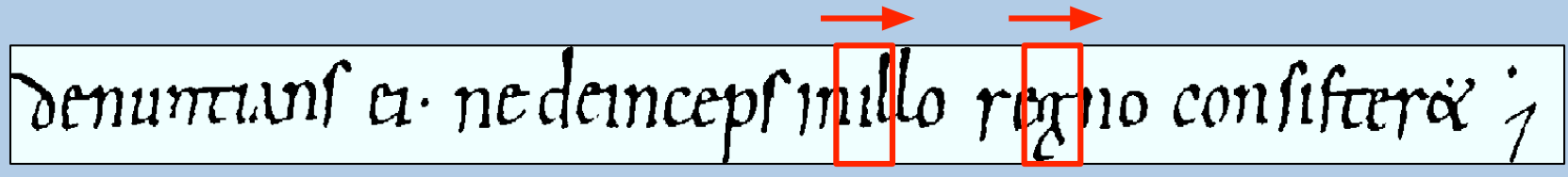


Outline

- Database
  - Ground Truth Creation
- Automatic Transcription
  - **Graph Similarity Features**
  - Fast Recognition Algorithm
- Keyword Spotting
- Transcription Alignment
- Conclusions

Standard Statistical Approach

- Fixed-size window, fixed-size vectorial features.
- Disregards pattern complexity and sub-part relations.



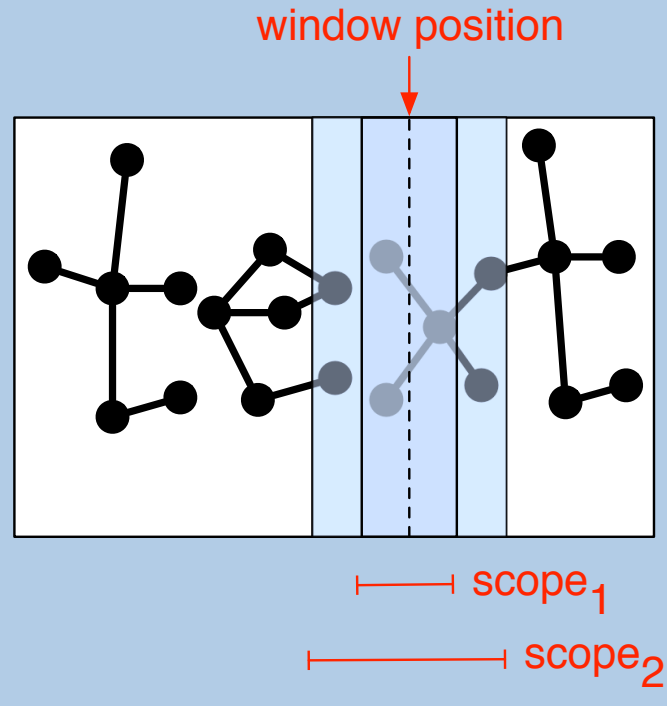
x_1 x_1

... ...

x_n x_n

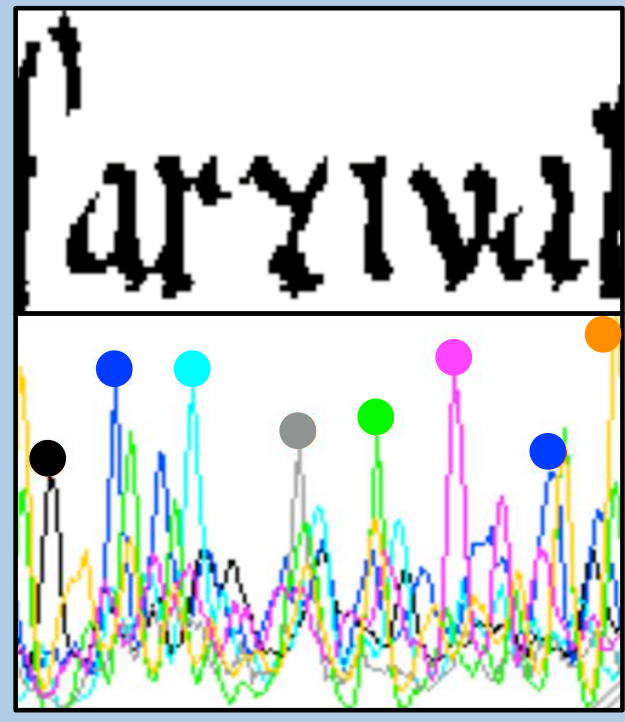
Proposed Structural Approach

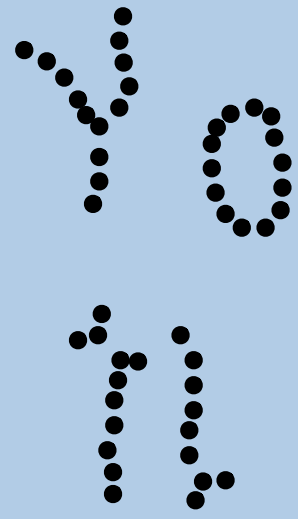
- Dynamic window size, graph-based handwriting representation.
- However: most basic mathematical operations not available for graphs.



Graph Similarity Features

- Bridge the gap with vector space embedding (Riesen & Bunke, 2010).
- Calculate character prototype similarity at each window position.





Word Recognition Results

- Significantly outperforms statistical features on the Parzival database.
- HMM error reduction with respect to (Marti & Bunke, 2001): 50%

$\rightarrow$ Enables statistical recognition of arbitrary structural representations.

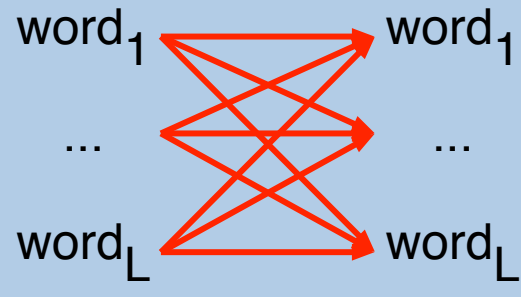
Features	Acc.
Marti & Bunke, 2001	88.69
Vinciarelli *et al.*, 2004	90.49
Fischer *et al.*, 2011	94.51

Outline

- Database
  - Ground Truth Creation
- Automatic Transcription
  - Graph Similarity Features
  - **Fast Recognition Algorithm**
- Keyword Spotting
- Transcription Alignment
- Conclusions

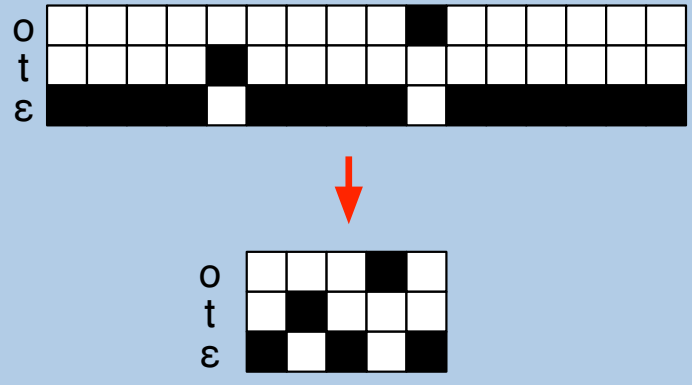
Computational Complexity

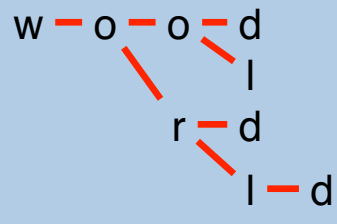
- Recognition: computational time $\propto$ word lexicon size^2
- *Minutes* per text line, *years* for the George Washington collection.

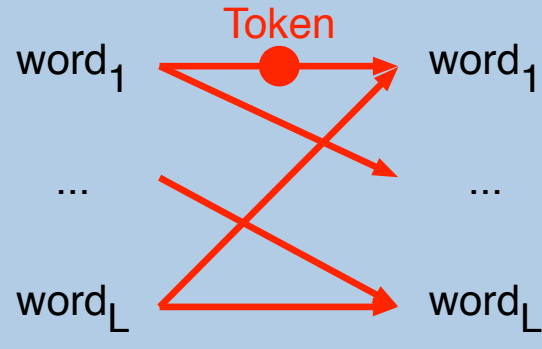


Search Space Optimization

- Compression of subsequent “no character” windows for NN.
- Word prefix based recognition, keep only few promising partial results.



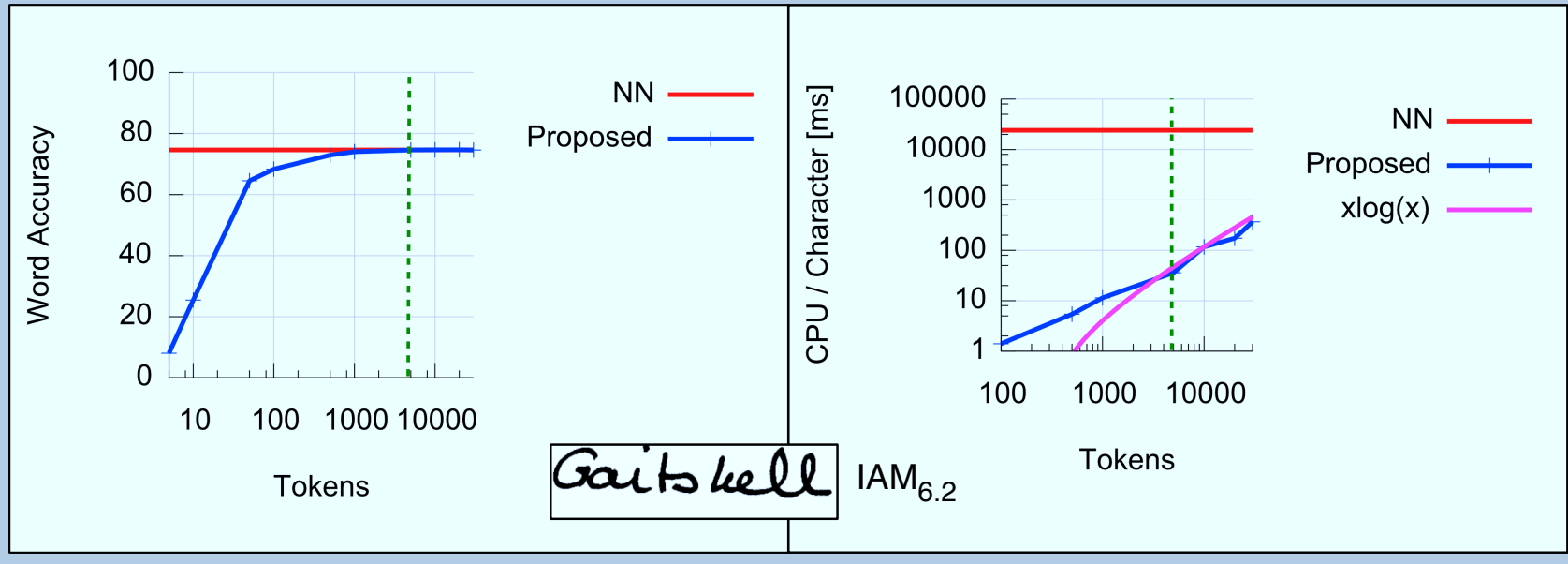




Speedup Results

- Over 600 times faster without accuracy loss for $20,000$ lexicon words.
- *Seconds* per text line, *weeks* for the George Washington collection.

→ One of the most accurate and fastest recognition systems worldwide.
→ Furthermore: versatile recognition output in form of word lattices.

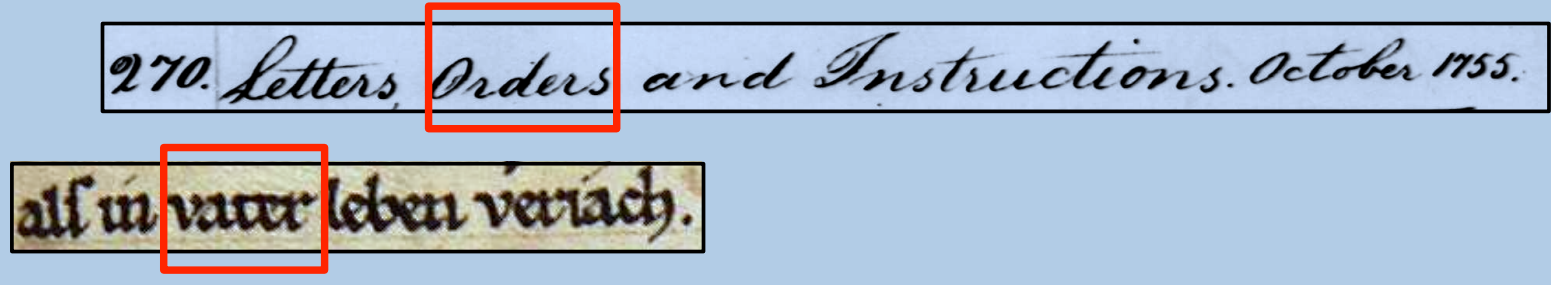


Outline

- Database
  - Ground Truth Creation
- Automatic Transcription
  - Graph Similarity Features
  - Fast Recognition Algorithm
- **Keyword Spotting**
- Transcription Alignment
- Conclusions

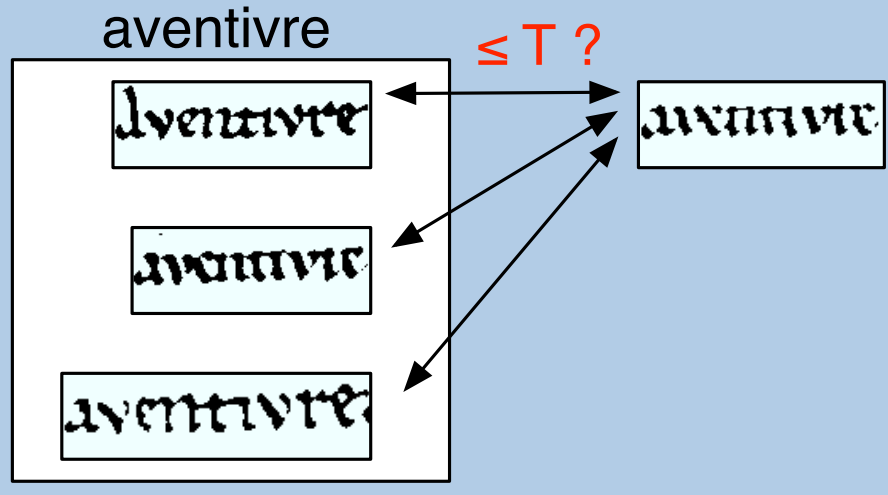
Keyword Spotting

- Identify search terms without transcription (Manmatha *et al.*, 1996).
- Lexicon-free indexing of manuscript images for digital libraries.



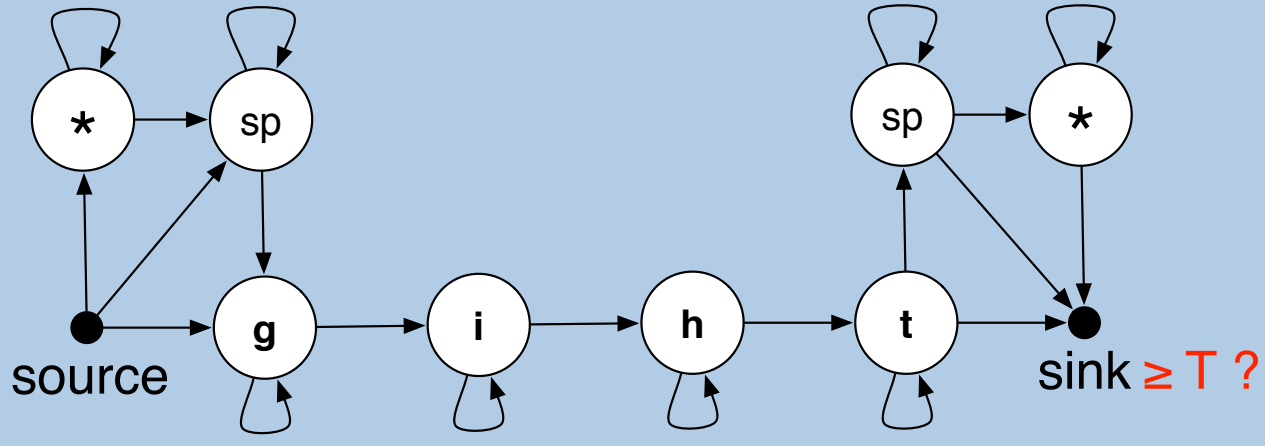
Standard Template-Based Approach

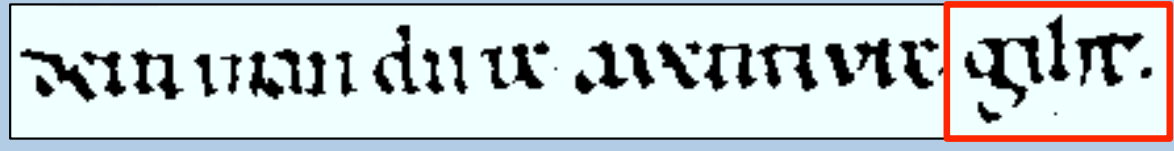
- Image matching between keyword templates and the manuscript.
- Dynamic Time Warping (DTW) distance $\leq$ threshold ?



Proposed Learning-Based Approach

- Concatenate trained character models to keywords.
- Recognition confidence at the most likely position $\geq$ threshold ?

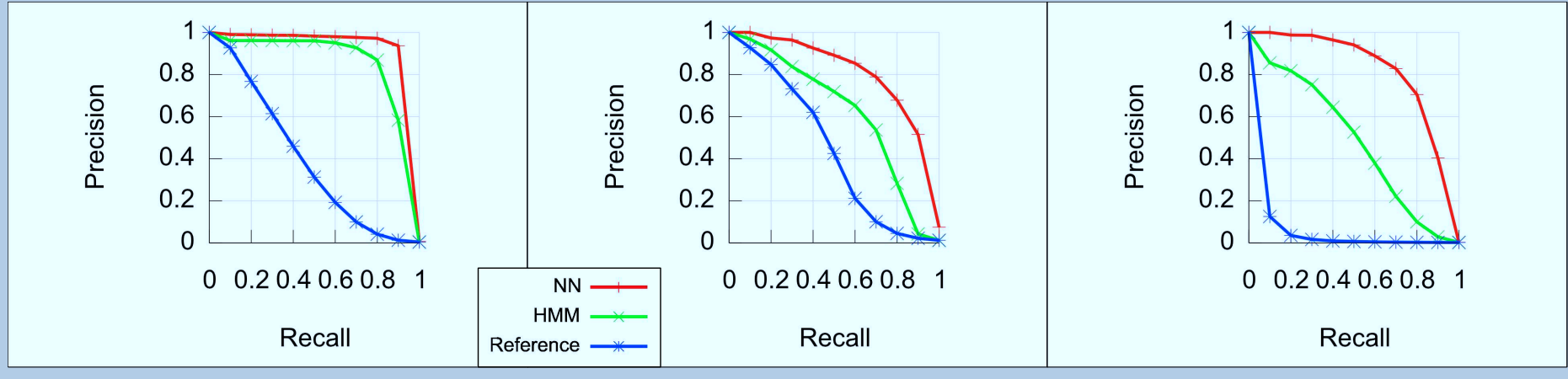


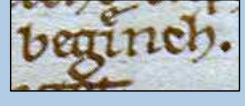


Spotting Results

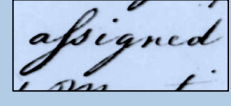
- NN average precision: 95% $\text{PAR}_{2.2}$, 81% $\text{GW}_{0.3}$, 83% $\text{IAM}_{6.2}$
- *Milliseconds* per text line, *hours* for the George Washington collection.

$\rightarrow$ Best performance on the George Washington database worldwide.

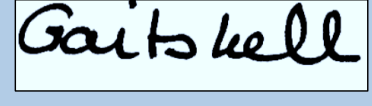




$\text{PAR}_{2.2}$



$\text{GW}_{0.3}$



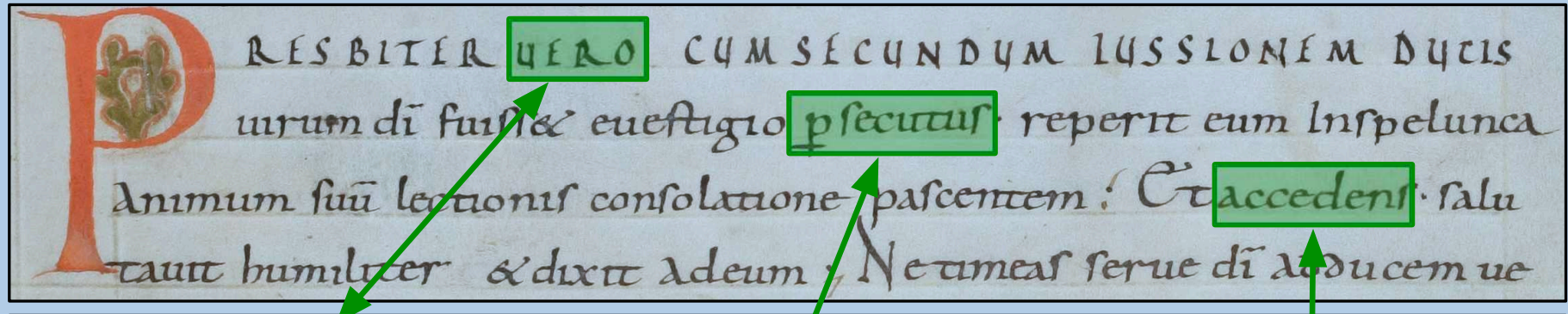
$\text{IAM}_{6.2}$

Outline

- Database
  - Ground Truth Creation
- Automatic Transcription
  - Graph Similarity Features
  - Fast Recognition Algorithm
- Keyword Spotting
- **Transcription Alignment**
- Conclusions

Transcription Alignment

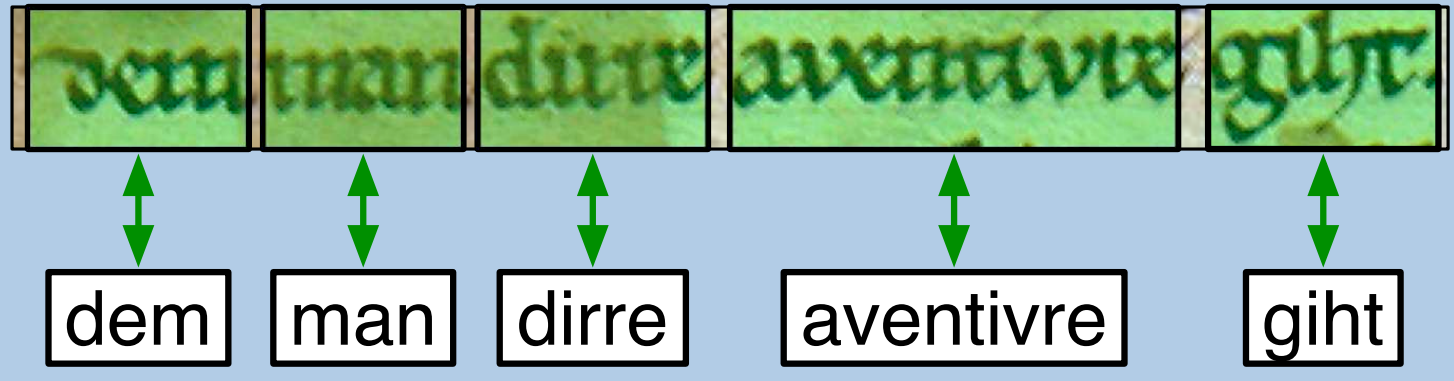
- Existing text editions often deviate from the manuscript image.
- Automatic alignment needed to extract training samples.



Presbyter vero cum, secundum iussionem ducis,
virum Dei fuisset e vestigio prosecutus, reperit eum in spelunca
animum suum lectionis consolatione pascentem; et accedens salu-
tavit humiliter, et dixit ad eum: Ne timeas, serve Dei, ad ducem ve-

Standard Segmentation Problem

- $1:1$ correlation between text and image.
- Segmentation of text line data sets (Zimmermann & Bunke, 2002).

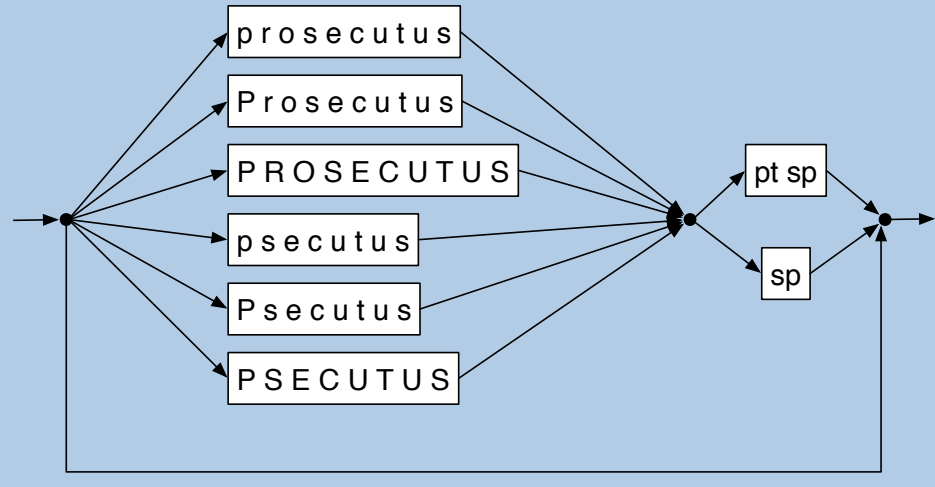


Inaccurate Transcriptions

- $n : m$ correlation between text and image.
- Images and texts without correspondence have to be rejected.

Proposed Alignment Approaches

- All text errors: multi-pass HMM recognition with a page lexicon.
- Spelling errors: HMM segmentation with spelling variants.



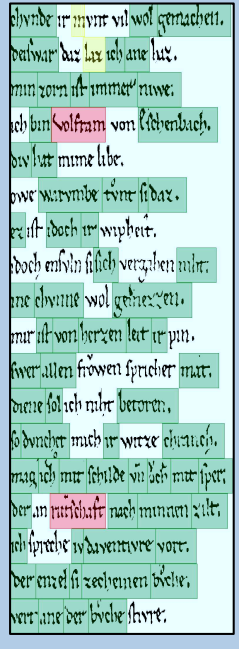
Transcription: dem man dirre aventivre giht.

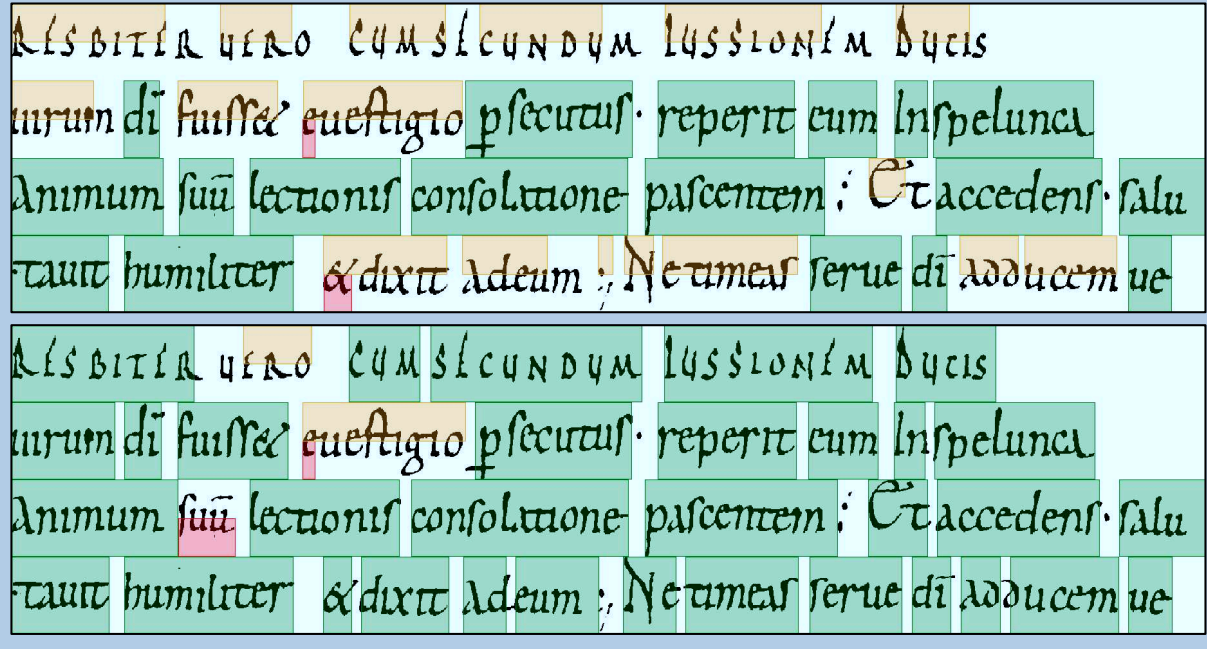
Edition: dem dirre diz wille giht.

Alignment Results

- 50% artificial text errors: 6% $PAR_{2.2}$ alignment error.
- Spelling errors: 8% $SG_{0.5}$, 17% $SG_{0.02}$ alignment error.

$\rightarrow$ Automatic database creation feasible with few initial learning samples.

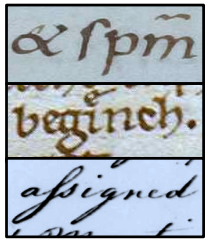
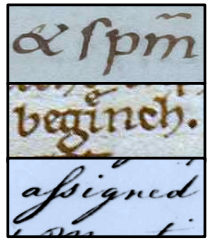
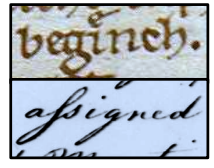
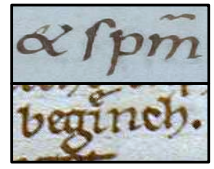
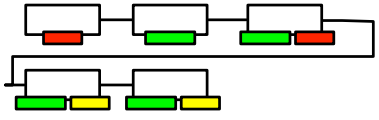
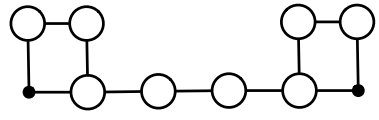
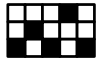




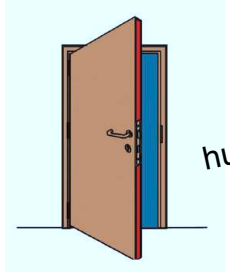
Outline

- Database
  - Ground Truth Creation
- Automatic Transcription
  - Graph Similarity Features
  - Fast Recognition Algorithm
- Keyword Spotting
- Transcription Alignment
- **Conclusions**

Contributions

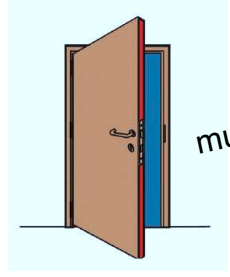
1 Database	2 Transcription	3 Spotting	4 Alignment
 $SG_{0.5}$ $PAR_{2.2}$ $GW_{0.3}$	 25% 6% 7% 18% N H	 83% 95% 81% N H	 8% 6% H
	 -50% error		
laypersons	 x 600 speed	> template-based	learning samples
10 minutes / page	seconds / line	milliseconds / line	milliseconds / line

Outlook



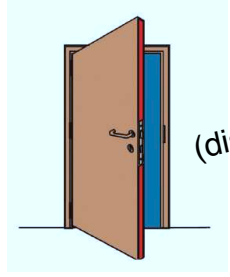
learn from human interaction

ground truth creation



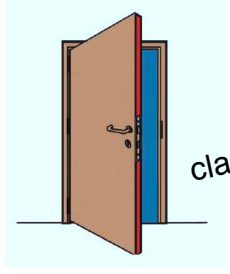
multiple scripts

single manuscript recognition



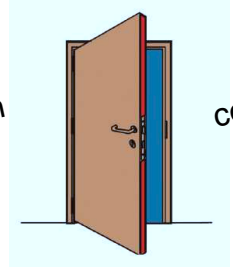
(dis)similarity measures

graph similarity features



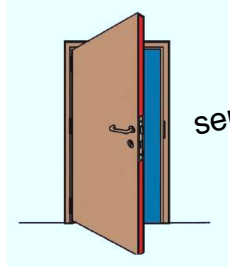
large lexicons
classifier combination

fast recognition algorithm



confidence modeling

keyword spotting



semi-supervised learning

transcription alignment

Not error-free, but accuracy & speed & costs are promising.

Questions