Interactive Approaches to Video Lecture Assessment

August 13, 2012

Korbinian Riedhammer

Speech Group
Pattern Recognition Lab

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Motivation

Computer Science 188, 001|Spring 2012|UC Berkeley

Pieter Abbeel

Description

COMPSCI 188 LEC 001 Sp12

▲	Name	Time	Released	Description	Popularity	Price
1	Lecture 8	--	2/10/12			FREE
2	Lecture 7	--	2/8/12			FREE
3	Lecture 6	--	2/2/12			FREE
4	Lecture 5	--	2/1/12			FREE
5	Lecture 2: audio only available	--	1/30/12			FREE
6	Lecture 4	--	1/26/12			FREE
7	Lecture 3	--	1/24/12			FREE
8	Lecture 1	--	1/18/12			FREE

Total: 8 Items

Subscribe Free

Get All

Category: Computer Science
Language: English
Copyright UC Regents 2012

More From UC Ber...

Journalism 102AC, 001|S...
Chemistry 3AL, 001|Sprin...
Environ Sci, Policy, and M...
Electrical Engineering 227...
Environ Sci, Policy, and M...

Links

Website
Report a Concern

Customer Ratings

We have not received enough ratings to display an average for this collection.

Click to rate

Customer Reviews

Be the first to write a review ›

Home > Courses > Electrical Engineering and Computer Science > Introduction to Computer Science and Programming > Video Lectures > 1: Introduction and Goals; Data Types, Operators, and Variables

1: Introduction and Goals; Data Types, Operators, and Variables

› VIEW ALL COURSES

› Course Home
› Syllabus
› Calendar
› Readings
› Video Lectures
› Assignments
› Exams

› Download Course Materials

Archived Versions
› Fall 2007

Join Study Group

Course Study Group

i have a program assignment. this is the problem: Input a number and store the integers of the number to the elements of the array. print th... 1 reply

Hello! I just started how to program in python and I typed -8%5 and the result was 2 instead of -3. Could someone explain this to me 4 replies

I am on lecture 8. I get

next »

About this Video | Playlist | Transcript | Download this Video

Topics covered: goals of the course; what is computation; introduction to data types, operators, and variables

Instructor: Prof. Eric Grimson, Prof. John Guttag

DONATE NOW

WHY I DONATE

Ashwini
Independent Learner
India

"Because I believe words, knowledge and ideas transform the world. By making OCW so accessible, you are creating free flow of knowledge."

Read more

Support OCW

Learn More >>

NTU VIRTUAL INSTRUCTOR

Lecture Search: SEARCH

INTRODUCTION TO DIGITAL SPEECH PROCESSING
Chapter 1
1-1
1-2
1-3
1-4
1-5
1-6
1-7
1-8
1-9
1-10
1-11
1-12
1-13
1-14

1-7 VOICE-BASED INFORMATION RETRIEVAL

LENGTH:
0:01:25.5

TIME SPAN OF THIS CHAPTER:
CH 1 CH 18

TIME SPAN OF THIS SLIDE:
1-1 1-27

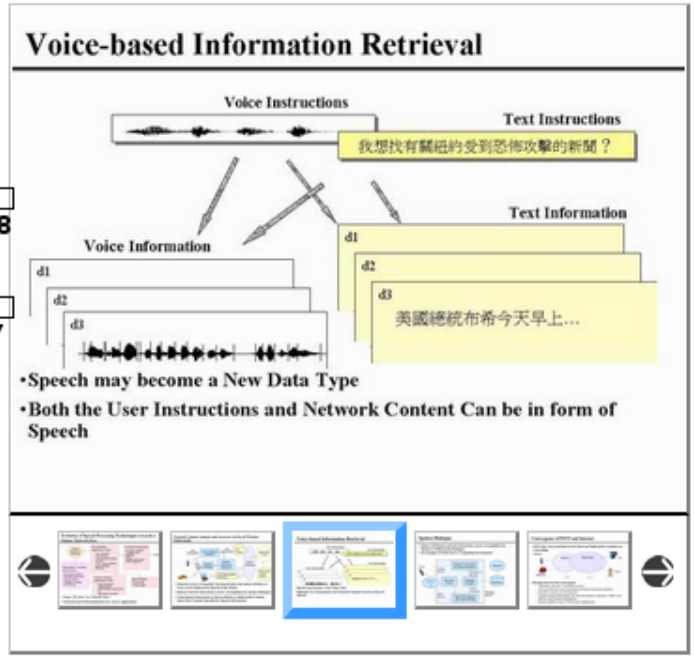
Play Summary (0:00:8.6)

Play Whole

KEY TERMS:

information retrieval
query
language model
adaptation
topic
special interest group

Related Key Terms



FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Videoportal

HERUNTERLADEN. ANSCHAUEN. LERNEN

SUPPORTED BY RRZE

VORLESUNG VERPASST ?
www.itunes.uni-erlangen.de

6 - Musteranalyse/Pattern Analysis

Zurück zur Clipansicht

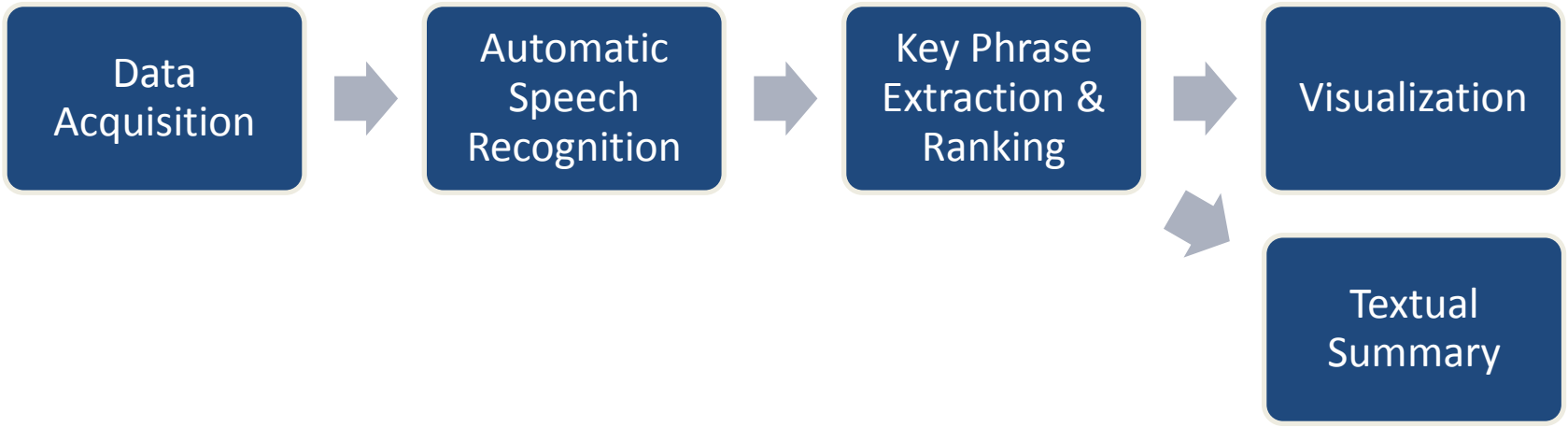
Nach Oben

Automatic key phrases

Visualization of the phrase occurrences

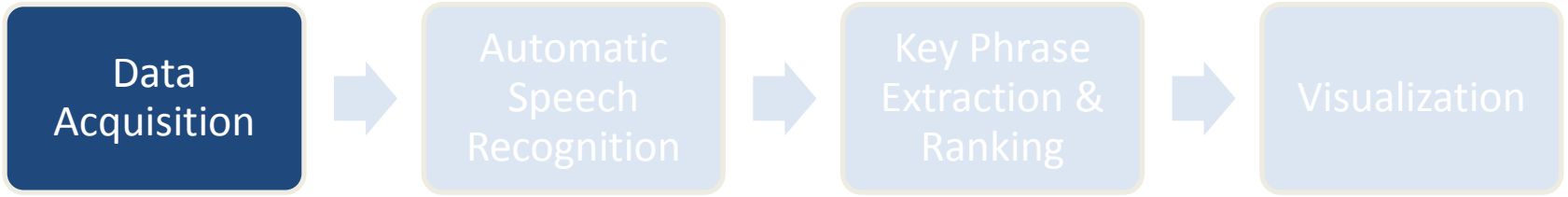
Search spoken text

Outline



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Data Acquisition



LMELectures – A Corpus of Academic Spoken English

- Two lecture series read in 2009
  - Pattern Analysis (PA)
  - Interventional Medical Image Processing (IMIP)
- 18 recordings per series
- About 40 hours of audio/video data
  - Audio: 48 kHz, 16 bit (AIFF), resampled to 16 kHz
  - Video: HD, reduced resolutions available due to bandwidth
- Clip-on cordless speaker microphone, room microphones
- Constant recording setting
  - RRZE E-Studio
  - Single speaker
  - Same recording equipment

Transcription

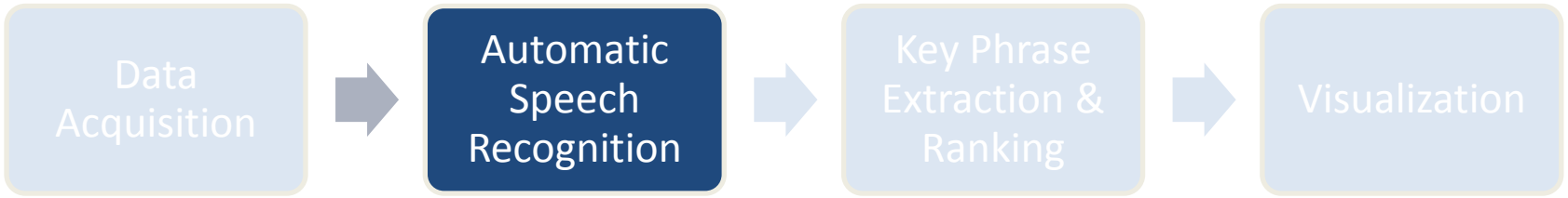
- Semi-automatic segmentation into speech turns
  - Based on speech pauses and silences
  - 23,857 turns
  - Average duration of 4.4 seconds
  - Total of about 29 hours of speech
- Manual transcription
  - New tool for the rapid transcription of speech
  - Time effort: about 5 times real time
- Transcription results
  - On average 14 words per speech turn
  - 300,500 words transcribed
  - Vocabulary size: 5,383 (excluding foreign words and word fragments)

Annotations

- Individual lecture *PA06*
- Based on edited manual transcript
- 5 human subjects
- 20 phrases
  - Ranking
  - Salience: from *1 (very relevant)* to *6 (useless)*

- Further annotations
  - Lecturer's key terms for series *PA*
  - Presentation slides in PDF format

Automatic Speech Recognition



The Kaldi Speech Recognition Toolkit

- State of the art, open source
- 4-layer system modeled by weighted finite state transducers (WFST)
  - Statistical n-gram language model
  - Lexicon with pronunciation alternatives
  - Context dependent phonemes
  - Hidden Markov models
- Acoustic frontend
  - Mel-frequency cepstral coefficients (MFCC), 1st and 2nd order derivatives
  - Phoneme dependent linear transformations
- Acoustic modeling: subspace Gaussian mixture models

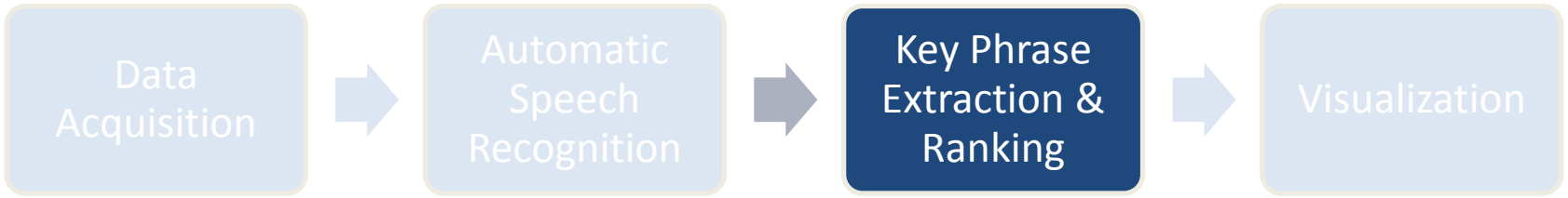
The *LMELectures* Speech Recognition System

- 600 Gaussian components, 5,500 HMM states
- Vocabulary size: 5,383
- Language model
  - 5,370,040 bi- and tri-grams
  - Trained on 500+ million words (including spontaneous lecture speech)

Name	Duration	# Turns	# Words	% WER
train	25h 31m	20,214	250,536	-
development	2h 07m	1,802	21,909	**9.78**
test	2h 12m	1,750	23,497	**11.03**

WER: word error rate

Key Phrase Extraction & Ranking



Candidate Selection

- A *verb alone* may be vague
  - "discuss" – what?
- An *isolated noun* may be ambiguous
  - "question" – difficult or easy?
- Information about the topic is often in the *noun phrase*
  - *"He asked a difficult question about the modified processing of words".*

- Apply part-of-speech tagging
- Extract noun phrases based on regular expression

Example

"it	*computes*	*the*	*principal*	*axes*	*that*	*'s*	*the*	*one*	*d*	*axis*		
PRP	VBZ	AR	ADJ	NN	IN	VBZ	AR	NUM	NN	NN		
that	*shows*	*the*	*highest*	*spread*	*of*	*the*	*points"*					
IN	VBZ	AR	ADJ	NN	AR	AR	NN					

adjective* (noun,number)$^+$ (article$^+$ adjective* (noun,number)$^+$)*

ADJ NN NUM AR ADJ NN NUM

Example

“it computes the principal axes that ‘s the one d axis that shows the highest spread of the points”

axes
principal axes
one
d
axis
one d
d axis
spread
highest spread
points
spread of the points
highest spread of the points

Unsupervised Ranking

- Frequent phrases may be salient
- With a similar occurrence count, longer phrases may be more salient
- Motivated by
  - *Didactics*: less confusion by literal repetition
  - *Psycholinguistics*: lexical entrainment

$$weight(phrase_i) = \begin{cases} f_i & , \quad n_i = 1 \\ f_i * (n_i + 1) & , \quad n_i > 1 \end{cases}$$

- Data and domain independent: simple and reliable
- Other investigated strategies include prior world or domain knowledge

Comparison of Rankings

- Compare a target ranking against a reference (human) ranking
- Standard measure: *Normalized Distributed Cumulative Gain*
- Award credit for placing valuable phrases at high ranks
- Compare lists of a certain length, e.g., top 10 phrases
- Phrases annotated with salience from *1 (very useful)* to *6 (useless)*

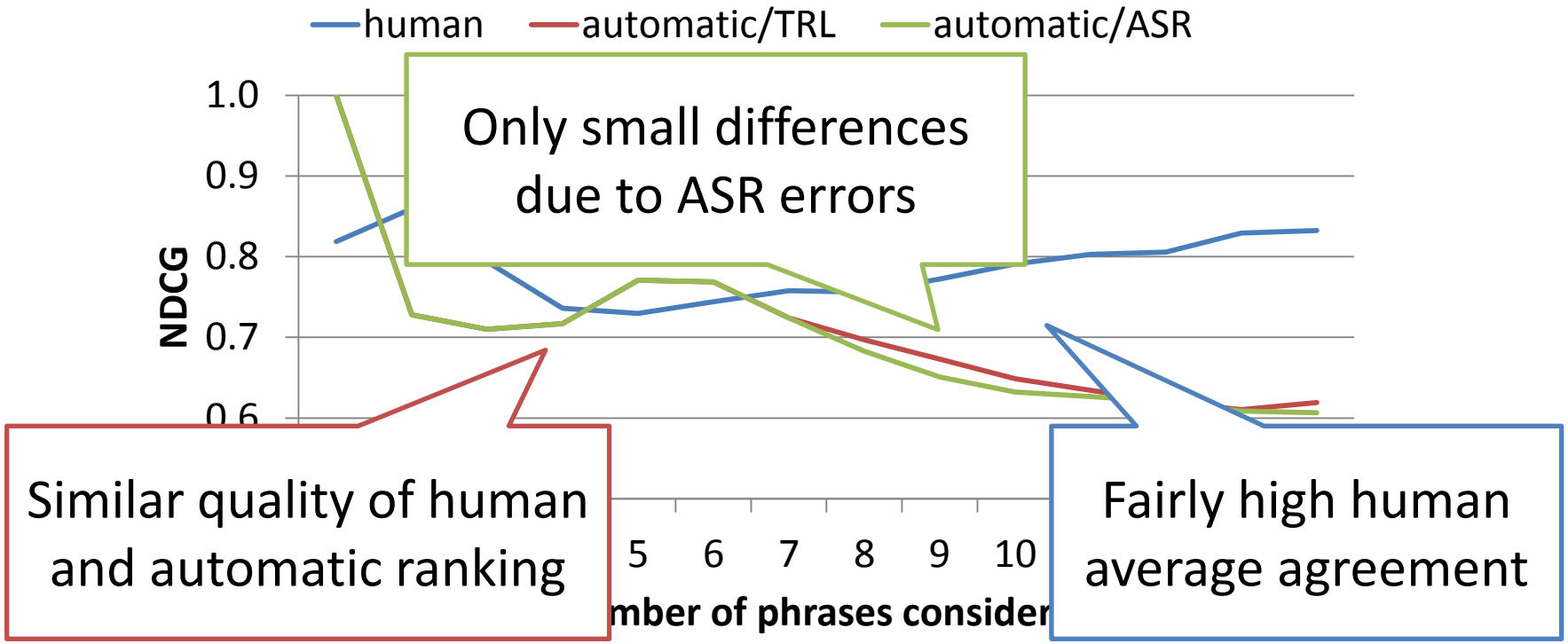
$$gain(phrase_i) = 2^{(6\,-grade_i)/5} - 1$$

$$NDCG(N) = C * \sum_{i=1}^{N} \frac{gain(phrase_i)}{\mathrm{ld}(1+i)}$$

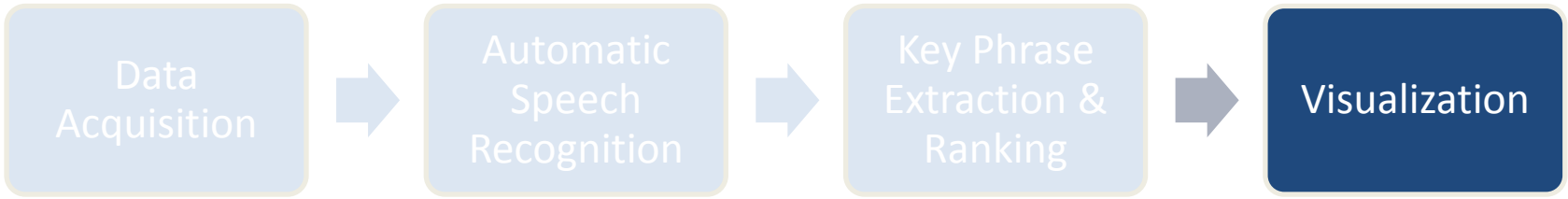
Multiple Annotators – Objective Results

- NDCG for pair-wise comparison only
- 5 human annotators
- *Human* score
  - average NDCG value of all human-human pairings
  - 20 individual pairings
- *Automatic* score
  - average NDCG value for all human-machine pairings
  - 5 individual pairings
- Scores based on manual *(TRL)* and automatic *(ASR)* transcripts

Evaluation of Human and Automatic Rankings



Visualization

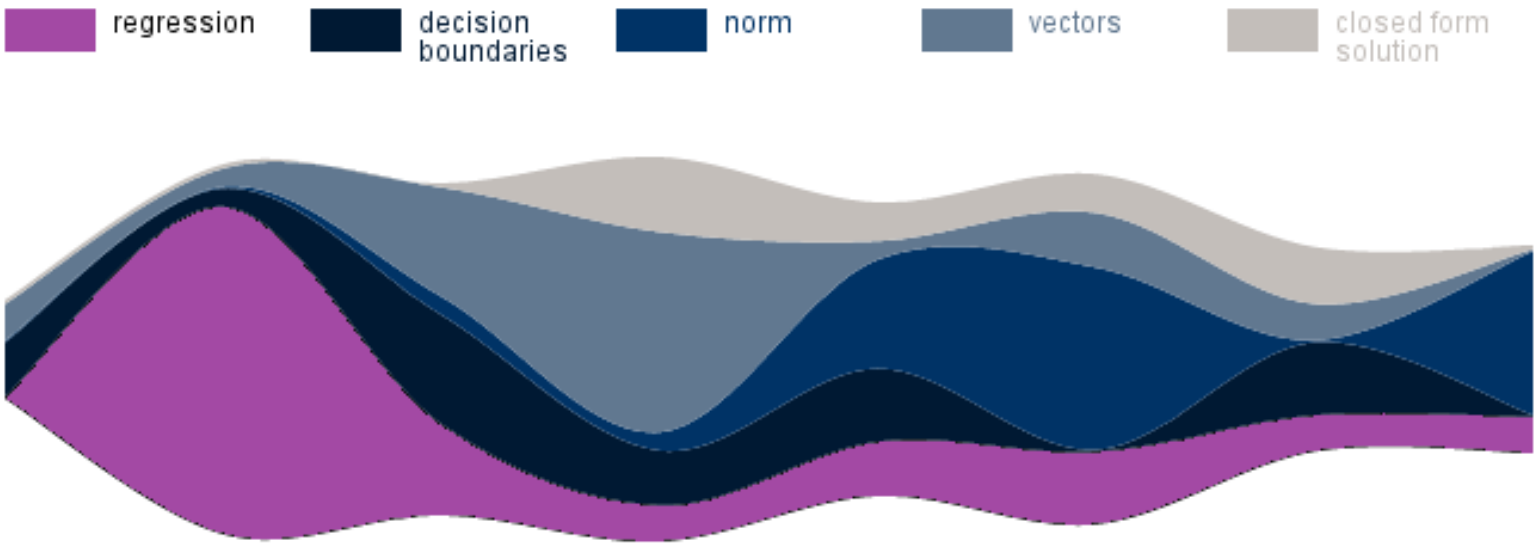


Motivation

- Key phrases give a topical overview of the lecture
- Phrase occurrences can serve as a visual index or navigation aid
- Simple example: clickable occurrence bar

StreamGraphs

- Popular in the visualization community
- Stacked splines
- Left to right: Playback time (as with occurrence bar)
- Stream wideness: Current phrase dominance
- Dominance: Number of occurrences within certain time frame



Advantages

- Comfortably display 3 to 6 phrases simultaneously
- Stream width can suggest topical relations of phrases
  - Similar widths at the same time indicate co-occurrence – possibly *related*
  - Different widths indicate rare or no co-occurrence – possibly *unrelated*

User Interactions

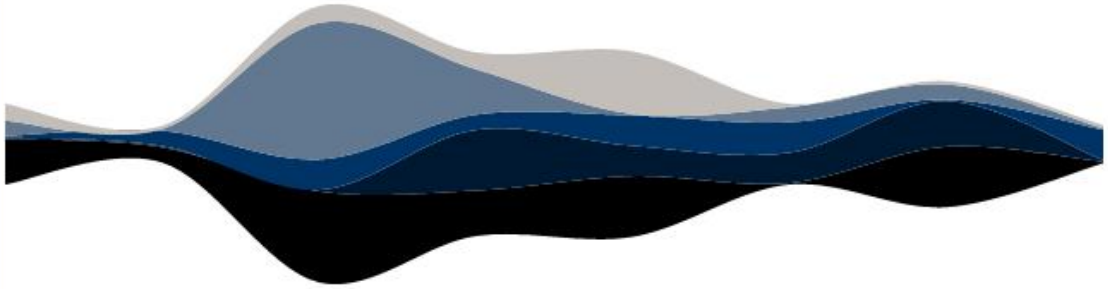
- Click into the stream – jump to closest occurrence
- Change the phrases on display – learn about topics and relations
- Interactions can be logged to collect data for customized rankings

Too long, didn't watch!

localhost/TIdw.html?Idescriptor=Idesc/20090512-Hornegger-PA06.xml

Musteranalyse/Pattern Analysis - 06

decision boundaries | closed form solution | linear regression | classes | linear decision boundaries



selected phrases

- decision boundaries
- closed form solution
- linear regression
- classes
- linear decision boundaries

search phrase

search phrase

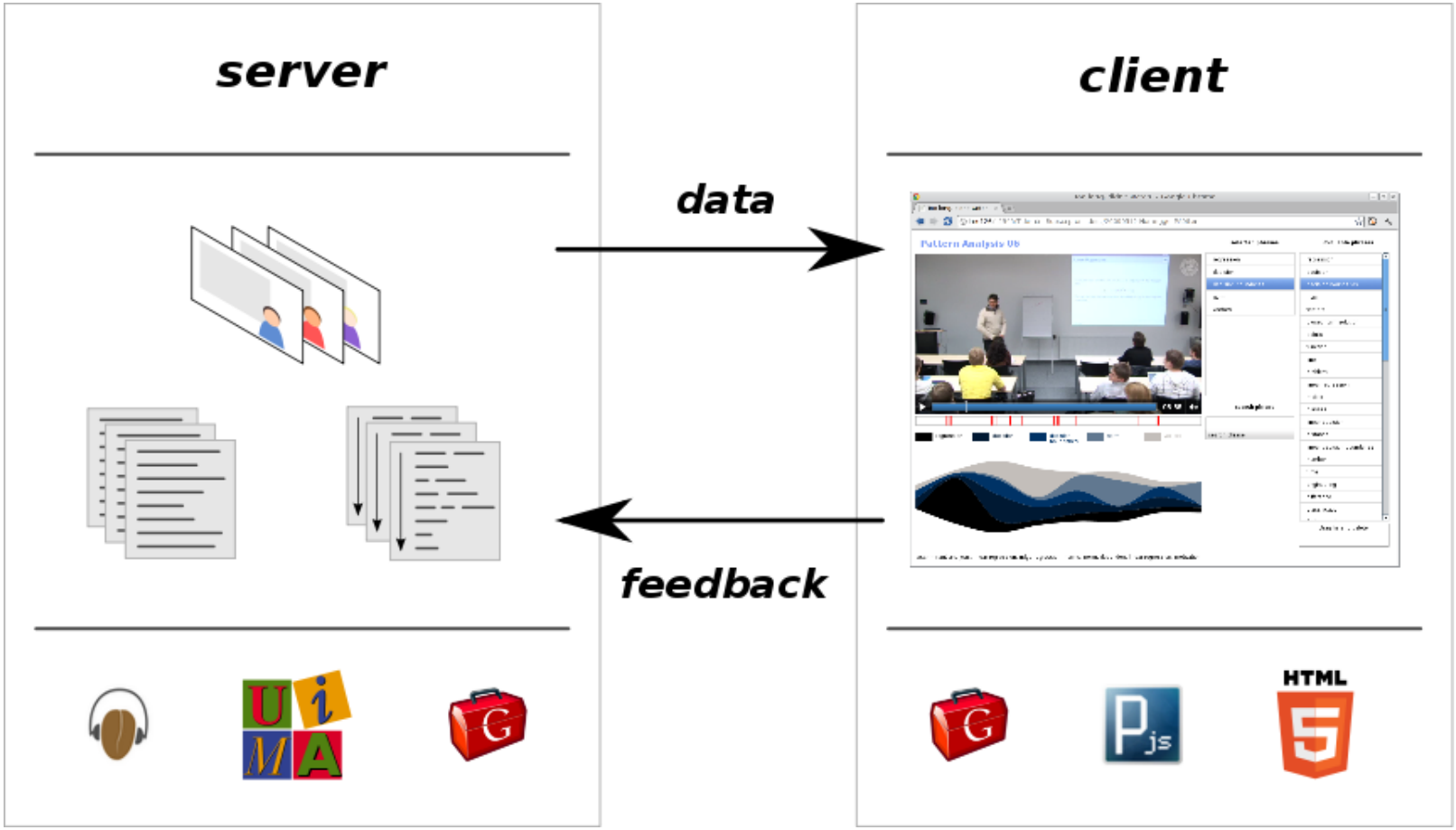
available phrases

- decision boundaries
- closed form solution
- linear regression
- classes
- linear decision boundaries
- norm
- mechanical engine
- ideal decision
- highest spread
- dependent linear regression
- engineering
- linear programming
- electrical engineering
- computational engineering
- modeling
- scenario
- addresses
- unit length
- straight line
- valued parameters

Drag here to delete.

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Implementation Details



User Study

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Task Based Evaluation

- Typical scenario: preparation for an exam
- Task should be independent of prior knowledge and comprehension
- Locate those segments of the video that cover certain topics
- Two groups of CS graduate students – *test* and *control*
  - Familiar with topic, speaker and lecture
  - 5 subjects per group
- Each participant is provided with
  - 3 lecture topics with short description
  - *Control* group: Video only
  - *Test* group: the presented interface
- Post-use questionnaire for *test* group to gather feedback

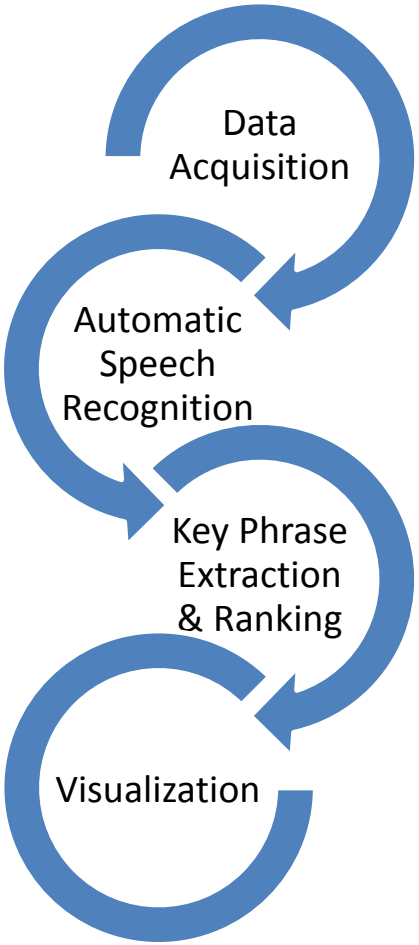
Results

Group	Accuracy	Average time
Control	68 %	30
Test	69 %	21

average time in minutes

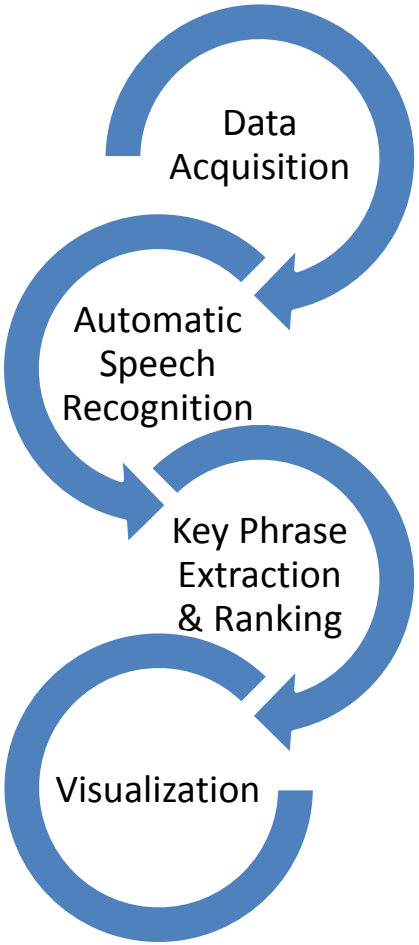
- Both groups have a similar accuracy
- Video duration: 42 minutes
- The *test* group was on average about 29% faster
- Most users found the…
  - interface to be helpful and easy to use
  - key phrase visualization to give a good overview

Summary



- *LMELectures*, a new corpus of academic spoken English
- Automatic speech recognition system for the *LMELectures* with a word error rate of 11%
- Unsupervised key phrase extraction and ranking that highly correlates to human rankings
- Novel video lecture browser that helps students to quickly assess the contents

Outlook



- More transcriptions for better acoustic and language models
- Integration of prior knowledge about speaker, room and topic
- Supervised methods for user-tailored rankings
- Larger user study on more lectures