

# Topics Seminar Media Informatics

Updated: December 21, 2017

Jörg Cassens

Seminar Medieninformatik

Winter Term 2017/2018



## Topics

A list of texts not assigned yet can be found on page 16.

### 1. Ambient Systems & Contextualised Computing

#### 1.1. Smart Meeting Rooms in Science and Education

☒ **Presented by: Lorenz Habenicht**

▷ Paper: AL Ronzhin, AA Karpov: A software system for the audiovisual monitoring of an intelligent meeting room in support of scientific and education activities. *Pattern Recognition and Image Analysis*, April 2015, Volume 25, Issue 2, pp 237-254. Springer

[Download paper](#)

*Abstract: This paper presents an analytical review of prototypes of intelligent spaces for scientific and education activities equipped with automatic recording tools and describes the specifics and distinguishing features of the intelligent meeting room developed at the St. Petersburg Institute for Informatics and Automation of RAS. A functional model of the automation of audiovisual monitoring of participants of activities is proposed based on the use of space-time data structuring describing the behavior of the participants within the analyzed room. Modern technologies of digital signal processing and pattern recognition have been used in implementing the proposed functional model of audiovisual monitoring in the smart conference room. New methods have been developed, in particular, the method of registration of participants of activities and the method of audiovisual recording of their presentations. The paper presents a software system for audiovisual monitoring for the automation of support for research and education activities held in a smart conference room. The main goal of the developed system is to identify events in the smart conference room, such as the time when a new user enters the room, when a speech begins, or when an audience member is given the floor. Experimental data on the participants were collected in the course of a simulation of activities where users held a meeting according to a given scenario and at real research and education activities when the participants were informed about the audiovisual recording of their behavior but it did not affect their planned activities in a smart conference room. During tests of the method of registration of participants of the event held in the smart conference room, more than 21000 photographs were taken. The average time required for taking a photograph of a participant was 1.3 s. The average displacement of the participant's face relative to the photographic center was 9%. The average person's face took up 30% of the area of the photograph. In addition, accumulated experimental data made it possible to identify places in the conference room from which questions were asked most frequently. The accuracy of pointing of the video camera at the speaker in the presentation area, as well as in rows of seats, was assessed by the size and position of their face in the frame over the entire process, averaging at 90%.*

#### 1.2. Human interaction with smart environments

☒ **Presented by: Eugen Kromm**

▷ Paper: AAN Shirehjini, A Semsar. Human interaction with IoT-based smart environments. *Multimedia Tools and Applications*, pp. 1-23. 2016. Springer

[Download paper](#)

**Abstract:** *This paper describes concepts, design, implementation, and performance evaluation of a 3D-based user interface for accessing IoT-based Smart Environments (IoT-SE). The generic interaction model of the described work addresses some major challenges of Human-IoT-SE-Interaction such as cognitive overload associated with manual device selection in complex IoT-SE, loss of user control, missing system image or over-automation. To address these challenges we propose a 3D-based mobile interface for mixed-initiative interaction in IoT-SE. The 3D visualization and 3D UI, acting as the central feature of the system, create a logical link between physical devices and their virtual representation on the end user's mobile devices. By so doing, the user can easily identify a device within the environment based on its position, orientation, and form, and access the identified devices through the 3D interface for direct manipulation within the scene. This overcomes the problem of manual device selection. In addition, the 3D visualization provides a system image for the IoT-SE, which supports users in understanding the ambience and things going on in it. Furthermore, the mobile interface allows users to control the amount and the way the IoT-SE automates the environment. For example, users can stop or postpone system triggered automatic actions, if they don't like or want them. Users also can remove a rule forever. By so doing, users can delete smart behaviors of their IoT-SE. This helps to overcome the automation challenges. In this paper, we present the design, implementation and evaluation of the proposed interaction system. We chose smart meeting rooms as the context for prototyping and evaluating our interaction concepts. However, the presented concepts and methods are generic and could be adapted to similar environments such as smart homes. We conducted a subjective usability evaluation (ISO-Norm 9241/110) with 16 users. All in one the study results indicate that the proposed 3D-User Interface achieved a good high score according to the ISO-Norm scores.*

### 1.3. Touch-less interactive augmented reality game

☒ **Presented by: Christian Spruck**

▷ Paper: Z Lv, A Halawani, S Feng, S Ur Réhman, H Li: Touch-less interactive augmented reality game on vision-based wearable device. *Personal and Ubiquitous Computing*, July 2015, Volume 19, Issue 3, pp 551–567. Springer

[Download paper](#)

**Abstract:** *There is an increasing interest in creating pervasive games based on emerging interaction technologies. In order to develop touch-less, interactive and augmented reality games on vision-based wearable device, a touch-less motion interaction technology is designed and evaluated in this work. Users interact with the augmented reality games with dynamic hands/feet gestures in front of the camera, which triggers the interaction event to interact with the virtual object in the scene. Three primitive augmented reality games with eleven dynamic gestures are developed based on the proposed touch-less interaction technology as proof. At last, a comparing evaluation is proposed to demonstrate the social acceptability and usability of the touch-less approach, running on a hybrid wearable framework or with Google Glass, as well as workload assessment, user's emotions and satisfaction.*

### 1.4. Privacy in Autonomous Systems

☒ **Presented by: Marco Di Maria**

▷ Paper: Nadin Kokciyan, Pinar Yolum: Context-Based Reasoning on Privacy in Internet of Things. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. AI and autonomy track*. Pages 4738-4744.

[Download paper](#)

**Abstract:** *More and more, devices around us are being connected to each other in the realm of Internet of Things (IoT). Their communication and especially collaboration promises useful services to be provided to end users. However, the same communication channels pose important privacy concerns to be raised. It is not clear which information will be shared with whom, for which intents, under which conditions. Existing approaches to privacy advocate policies to be in place to regulate privacy. However, the scale and heterogeneity of the IoT entities make it infeasible to maintain policies among each and every entity in the system. Conversely, it is best if each entity can reason on the privacy using norms and context autonomously. Accordingly, this paper proposes an approach where each entity finds out which contexts it is in based on information it gathers from other entities in the system. The proposed approach uses argumentation to enable IoT entities to reason about their context and decide to reveal information based on it. We demonstrate the applicability of the approach over an IoT scenario.*

▷ Paper: Jose M. Such: Privacy and Autonomous Systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. AI and autonomy track*. Pages 4761-4767.

[Download paper](#)

**Abstract:** *We discuss the problem of privacy in autonomous systems, introducing different conceptualizations and perspectives on privacy to assess the threats that autonomous systems may pose to privacy. After this, we outline socio-technical and legal measures that should be put in place to mitigate these threats. Beyond privacy threats and countermeasures, we also argue how autonomous systems may be, at*

the same time, the key to address some of the most challenging and pressing privacy problems nowadays and in the near future.

## 2. Natural Language: Image Description, Conversation, Storytelling

### 2.1. Survey of Automatic Description Generation

☒ **Presented by: Oxana Lupashko**

- ▷ Paper: Raffaella Bernardi, Ruket Cakici, Desmond Elliott, Aykut Erdem, Erkut Erdem, Nazli Ikizler-Cinbis, Frank Keller, Adrian Muscat, and Barbara Plank. 2016. Automatic description generation from images: a survey of models, datasets, and evaluation measures. *J. Artif. Int. Res.* 55, 1 (January 2016), 409-442.

[Download paper](#)

*Abstract: Automatic description generation from natural images is a challenging problem that has recently received a large amount of interest from the computer vision and natural language processing communities. In this survey, we classify the existing approaches based on how they conceptualize this problem, viz., models that cast description as either generation problem or as a retrieval problem over a visual or multimodal representational space. We provide a detailed review of existing models, highlighting their advantages and disadvantages. Moreover, we give an overview of the benchmark image datasets and the evaluation measures that have been developed to assess the quality of machine-generated image descriptions. Finally we extrapolate future directions in the area of automatic image description generation.*

### 2.2. Conversational Systems

☒ **Presented by: Junaid Ahmed Ghauri**

- ▷ Paper: Zhou Yu, Alexander Rudnicky, Alan Black. Learning Conversational Systems that Interleave Task and Non-Task Content. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main track. Pages 4214-4220

[Download paper](#)

*Abstract: Task-oriented dialog systems have been applied in various tasks, such as automated personal assistants, customer service providers and tutors. These systems work well when users have clear and explicit intentions that are well-aligned to the systems' capabilities. However, they fail if users intentions are not explicit. To address this shortcoming, we propose a framework to interleave non-task content (i.e. everyday social conversation) into task conversations. When the task content fails, the system can still keep the user engaged with the non-task content. We trained a policy using reinforcement learning algorithms to promote long-turn conversation coherence and consistency, so that the system can have smooth transitions between task and non-task content. To test the effectiveness of the proposed framework, we developed a movie promotion dialog system. Experiments with human users indicate that a system that interleaves social and task content achieves a better task success rate and is also rated as more engaging compared to a pure task-oriented system.*

- ▷ Paper: Oscar J. Romero, Ran Zhao, Justine Cassell: Cognitive-Inspired Conversational-Strategy Reasoner for Socially-Aware Agents. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Main track. Pages 3807-3813.

[Download paper](#)

*Abstract: In this work we propose a novel module for a dialogue system that allows a conversational agent to utter phrases that do not just meet the system's task intentions, but also work towards achieving the system's social intentions. The module - a Social Reasoner - takes the task goals the system must achieve and decides the appropriate conversational style and strategy with which the dialogue system describes the information the user desires so as to boost the strength of the relationship between the user and system (rapport), and therefore the user's engagement and willingness to divulge the information the agent needs to efficiently and effectively achieve the user's goals. Our Social Reasoner is inspired both by analysis of empirical data of friends and stranger dyads engaged in a task, and by prior literature in fields as diverse as reasoning processes in cognitive and social psychology, decision-making, sociolinguistics and conversational analysis. Our experiments demonstrated that, when using the Social Reasoner in a Dialogue System, the rapport level between the user and system increases in more than 35% in comparison with those cases where no Social Reasoner is used.*

### 2.3. Interactive Storytelling

☒ **Presented by: Kivanc Yilmaz**

- ▷ Paper: Reid Swanson and Andrew S. Gordon. 2012. Say Anything: Using Textual Case-Based Reasoning to Enable Open-Domain Interactive Storytelling. *ACM Transactions on Intelligent Interactive Systems*, 2(3).

[Download paper](#)

*Abstract: We describe Say Anything, a new interactive storytelling system that collaboratively writes*

textual narratives with human users. Unlike previous attempts, this interactive storytelling system places no restrictions on the content or direction of the user's contribution to the emerging storyline. In response to these contributions, the computer continues the storyline with narration that is both coherent and entertaining. This capacity for open-domain interactive storytelling is enabled by an extremely large repository of nonfiction personal stories, which is used as a knowledge base in a case-based reasoning architecture. In this article, we describe the three main components of our case-based reasoning approach: a million-item corpus of personal stories mined from internet weblogs, a case retrieval strategy that is optimized for narrative coherence, and an adaptation strategy that ensures that repurposed sentences from the case base are appropriate for the user's emerging fiction. We describe a series of evaluations of the system's ability to produce coherent and entertaining stories, and we compare these narratives with single-author stories posted to internet weblogs.

## 2.4. Cross-Lingual Event Tracking

□ **Topic not assigned**

- ▷ Paper: Jan Rupnik, Andrej Muhič, Gregor Leban, Primož Škraba, Blaž Fortuna, and Marko Grobelnik. 2016. News across languages – cross-lingual document similarity and event tracking. *J. Artif. Int. Res.* 55, 1 (January 2016), 283-316.

[Download paper](#)

*Abstract: In today's world, we follow news which is distributed globally. Significant events are reported by different sources and in different languages. In this work, we address the problem of tracking of events in a large multilingual stream. Within a recently developed system Event Registry we examine two aspects of this problem: how to compare articles in different languages and how to link collections of articles in different languages which refer to the same event. Taking a multilingual stream and clusters of articles from each language, we compare different cross-lingual document similarity measures based on Wikipedia. This allows us to compute the similarity of any two articles regardless of language. Building on previous work, we show there are methods which scale well and can compute a meaningful similarity between articles from languages with little or no direct overlap in the training data. Using this capability, we then propose an approach to link clusters of articles across languages which represent the same event. We provide an extensive evaluation of the system as a whole, as well as an evaluation of the quality and robustness of the similarity measure and the linking algorithm.*

## 3. Natural Language: Sentiment Analysis

### 3.1. Survey: Sentiment Analysis

□ **Topic not assigned**

- ▷ Paper: Maite Taboada. Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics* 2016 2:1, 325-347

[Download paper](#)

*Abstract: Sentiment analysis is a growing field at the intersection of linguistics and computer science that attempts to automatically determine the sentiment contained in text. Sentiment can be characterized as positive or negative evaluation expressed through language. Common applications of sentiment analysis include the automatic determination of whether a review posted online (of a movie, a book, or a consumer product) is positive or negative toward the item being reviewed. Sentiment analysis is now a common tool in the repertoire of social media analysis carried out by companies, marketers, and political analysts. Research on sentiment analysis extracts information from positive and negative words in text, from the context of those words, and from the linguistic structure of the text. This brief review examines in particular the contributions that linguistic knowledge can make to the task of automatically determining sentiment.*

### 3.2. Taxonomy of Negativity

☒ **Presented by: Mehran Amiri**

- ▷ Paper: Taboada, M., Trnavac, R. & Goddard, C.: On Being Negative. *Corpus Pragmatics* (2017) 1: 57-76.

[Download paper](#)

*Abstract: This paper investigates the pragmatic expressions of negative evaluation (negativity) in two corpora: (i) comments posted online in response to newspaper opinion articles; and (ii) online reviews of movies, books and consumer products. We propose a taxonomy of linguistic resources that are deployed in the expression of negativity, with two broad groups at the top level of the taxonomy: resources from the lexicogrammar or from discourse semantics. We propose that rhetorical figures can be considered part of the discourse semantic resources used in the expression of negativity. Using our taxonomy as starting point, we carry out a corpus analysis, and focus on three phenomena: adverb + adjective combinations; rhetorical questions; and rhetorical figures. Although the analysis in this paper is corpus-assisted rather than corpus-driven, the final goal of our research is to make it quantitative, in extracting patterns and*

resources that can be detected automatically.

### 3.3. Negation and Speculation

☒ **Presented by: Shayan Jawed**

- ▷ Paper: Cruz, N. P., Taboada, M. and Mitkov, R. (2016), A machine-learning approach to negation and speculation detection for sentiment analysis. *J Assn Inf Sci Tec*, 67: 2118–2136.

[Download paper](#)

Abstract: *Recognizing negative and speculative information is highly relevant for sentiment analysis. This paper presents a machine-learning approach to automatically detect this kind of information in the review domain. The resulting system works in two steps: in the first pass, negation/speculation cues are identified, and in the second phase the full scope of these cues is determined. The system is trained and evaluated on the Simon Fraser University Review corpus, which is extensively used in opinion mining. The results show how the proposed method outstrips the baseline by as much as roughly 20% in the negation cue detection and around 13% in the scope recognition, both in terms of F1. In speculation, the performance obtained in the cue prediction phase is close to that obtained by a human rater carrying out the same task. In the scope detection, the results are also promising and represent a substantial improvement on the baseline (up by roughly 10%). A detailed error analysis is also provided. The extrinsic evaluation shows that the correct identification of cues and scopes is vital for the task of sentiment analysis.*

### 3.4. Predicting Quality and Interestingness of Texts

□ **Topic not assigned**

- ▷ Paper: Debasis Ganguly, Johannes Leveling, and Gareth J.F. Jones. Automatic prediction of text aesthetics and interestingness. In *Proceedings of the International Conference on Computational Linguistics*. 2014.

[Download paper](#)

Abstract: *This paper investigates the problem of automated text aesthetics prediction. The availability of user generated content and ratings, e.g. Flickr, has induced research in aesthetics prediction for non-text domains, particularly for photographic images. This problem, however, has yet not been explored for the text domain. Due to the very subjective nature of text aesthetics, it is difficult to compile human annotated data by methods such as crowd sourcing with a fair degree of inter-annotator agreement. The availability of the Kindle “popular highlights” data has motivated us to compile a dataset comprised of human annotated aesthetically pleasing and interesting text passages. We then undertake a supervised classification approach to predict text aesthetics by constructing real-valued feature vectors from each text passage. In particular, the features that we use for this classification task are word length, repetitions, polarity, part-of-speech, semantic distances; and topic generality and diversity. A traditional binary classification approach is not effective in this case because non-highlighted passages surrounding the highlighted ones do not necessarily represent the other extreme of unpleasant quality text. Due to the absence of real negative class samples, we employ the MC algorithm, in which training can be initiated with instances only from the positive class. On each successive iteration the algorithm selects new strong negative samples from the unlabeled class and retrains itself. The results show that the mapping convergence (MC) algorithm with a Gaussian and a linear kernel used for the mapping and convergence phases, respectively, yields the best results, achieving satisfactory accuracy, precision and recall values of about 74%, 42% and 54% respectively.*

- ▷ Paper: Tong Wang, Ping Chen, Boyang Li: Predicting the Quality of Short Narratives from Social Media. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. Main track. Pages 3859-3865.

[Download paper](#)

Abstract: *An important and difficult challenge in building computational models for narratives is the automatic evaluation of narrative quality. Quality evaluation connects narrative understanding and generation as generation systems need to evaluate their own products. To circumvent difficulties in acquiring annotations, we employ upvotes in social media as an approximate measure for story quality. We collected 54,484 answers from a crowd-powered question-and-answer website, Quora, and then used active learning to build a classifier that labeled 28,320 answers as stories. To predict the number of upvotes without the use of social network features, we create neural networks that model textual regions and the interdependence among regions, which serve as strong benchmarks for future research. To our best knowledge, this is the first large-scale study for automatic evaluation of narrative quality.*

## 4. Summarization

### 4.1. Lecture Interaction & Summarization

☒ **Presented by: Alda Cypi**

- ▷ Paper: X Fan, W Luo, M Menekse, D Litman: CourseMIRROR: Enhancing large classroom instructor-student interactions via mobile interfaces and natural language processing, 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, Pages 1473-1478, 2015. ACM

[Download paper](#)

Abstract: *Interactions between students and instructors are crucial to the success of learning and teaching. However, such interactions are limited in large classrooms (e.g. STEM courses and MOOCs). We present CourseMIRROR (Mobile In-situ Reflections and Review with Optimized Rubrics), a mobile system that prompts students' self-reflection and in-situ feedback to enhance the interactions. CourseMIRROR uses automatic text summarization techniques to aggregate students' feedback and present the most significant ones to both the instructors and the students to help them understand both difficulties and misunderstandings encountered. In two semester-long pilot deployments involving 20 participants, we received positive feedback from both students and instructors. We highlight major findings as well as the lessons learned.*

- ▷ Paper: W Luo, D Litman: Summarizing student responses to reflection prompts. Proceedings of the 2015 Conference on Empirical Methods in NLP, pp. 1955-1960, Association for Computational Linguistics. 2015

[Download paper](#)

Abstract: *We propose to automatically summarize student responses to reflection prompts and introduce a novel summarization algorithm that differs from traditional methods in several ways. First, since the linguistic units of student inputs range from single words to multiple sentences, our summaries are created from extracted phrases rather than from sentences. Second, the phrase summarization algorithm ranks the phrases by the number of students who semantically mention a phrase in a summary. Experimental results show that the proposed phrase summarization approach achieves significantly better summarization performance on an engineering course corpus in terms of ROUGE scores when compared to other summarization methods, including MEAD, LexRank and MMR.*

- ▷ Paper: W Luo, X Fan, M Menekse, J Wang: W Luo, X Fan, M Menekse, J Wang, D Litman: Enhancing instructor-student and student-student interactions with mobile interfaces and summarization. Proceedings of NAACL-HLT 2015. Association for Computational Linguistics. 2015

[Download paper](#)

Abstract: *Educational research has demonstrated that asking students to respond to reflection prompts can increase interaction between instructors and students, which in turn can improve both teaching and learning especially in large classrooms. However, administering an instructor's prompts, collecting the students' responses, and summarizing these responses for both instructors and students is challenging and expensive. To address these challenges, we have developed an application called CourseMIRROR (Mobile In-situ Reflections and Review with Optimized Rubrics). CourseMIRROR uses a mobile interface to administer prompts and collect reflective responses for a set of instructor-assigned course lectures. After collection, CourseMIRROR automatically summarizes the reflections with an extractive phrase summarization method, using a clustering algorithm to rank extracted phrases by student coverage. Finally, CourseMIRROR presents the phrase summary to both instructors and students to help them understand the difficulties and misunderstandings encountered.*

### 4.2. Headline Generation

□ **Topic not assigned**

- ▷ Paper: Enrique Alfonseca, Daniele Pighin, and Guillermo Garrido. HEADY: news headline abstraction through event pattern clustering. In ACL 2013, pages 1243–1253, 2013.

[Download paper](#)

Abstract: *This paper presents HEADY: a novel, abstractive approach for headline generation from news collections. From a web-scale corpus of English news, we mine syntactic patterns that a Noisy-OR model generalizes into event descriptions. At inference time, we query the model with the patterns observed in an unseen news collection, identify the event that better captures the gist of the collection and retrieve the most appropriate pattern to generate a headline. HEADY improves over a state-of-the-art open-domain title abstraction method, bridging half of the gap that separates it from extractive methods using human-generated titles in manual evaluations, and performs comparably to human-generated headlines as evaluated with ROUGE.*

- ▷ Paper: Jiwei Tan, Xiaojun Wan, Jianguo Xiao: From Neural Sentence Summarization to Headline Generation: A Coarse-to-Fine Approach. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Main track. Pages 4109-4115.

### [Download paper](#)

*Abstract: Headline generation is a task of abstractive text summarization, and previously suffers from the immaturity of natural language generation techniques. Recent success of neural sentence summarization models shows the capacity of generating informative, fluent headlines conditioned on selected recapitulative sentences. In this paper, we investigate the extension of sentence summarization models to the document headline generation task. The challenge is that extending the sentence summarization model to consider more document information will mostly confuse the model and hurt the performance. In this paper, we propose a coarse-to-fine approach, which first identifies the important sentences of a document using document summarization techniques, and then exploits a multi-sentence summarization model with hierarchical attention to leverage the important sentences for headline generation. Experimental results on a large real dataset demonstrate the proposed approach significantly improves the performance of neural sentence summarization models on the headline generation task.*

#### 4.3. Time Aware Knowledge Extraction for Summarization

□ **Topic not assigned**

- ▷ Paper: Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, Mimmo Parente, Time Aware Knowledge Extraction for microblog summarization on Twitter, In Information Fusion, Volume 28, 2016, Pages 60-74, ISSN 1566-2535

### [Download paper](#)

*Abstract: Microblogging services like Twitter and Facebook collect millions of user generated content every moment about trending news, occurring events, and so on. Nevertheless, it is really a nightmare to find information of interest through the huge amount of available posts that are often noisy and redundant. In the era of Big Data, social media analytics services have caught increasing attention from both research and industry. Specifically, the dynamic context of microblogging requires to manage not only meaning of information but also the evolution of knowledge over the timeline. This work defines Time Aware Knowledge Extraction (briefly TAKE) methodology that relies on temporal extension of Fuzzy Formal Concept Analysis. In particular, a microblog summarization algorithm has been defined filtering the concepts organized by TAKE in a time-dependent hierarchy. The algorithm addresses topic-based summarization on Twitter. Besides considering the timing of the concepts, another distinguishing feature of the proposed microblog summarization framework is the possibility to have more or less detailed summary, according to the user's needs, with good levels of quality and completeness as highlighted in the experimental results.*

#### 4.4. Annotator Rationales for Relevance Judgment

□ **Topic not assigned**

- ▷ Paper: Tyler McDonnell, Matthew Lease, Mucahid Kutlu, and Tamer Elsayed. Why Is That Relevant? Collecting Annotator Rationales for Relevance Judgments. In Proc. of the 4th AAI Conference on Human Computation and Crowdsourcing (HCOMP), pages 139–148, 2016.

### [Download paper](#)

*Abstract: When collecting subjective human ratings of items, it can be difficult to measure and enforce data quality due to task subjectivity and lack of insight into how judges' arrive at each rating decision. To address this, we propose requiring judges to provide a specific type of rationale underlying each rating decision. We evaluate this approach in the domain of Information Retrieval, where human judges rate the relevance of Webpages to search queries. Cost-benefit analysis over 10,000 judgments collected on Mechanical Turk suggests a win-win: experienced crowd workers provide rationales with almost no increase in task completion time while providing a multitude of further benefits, including more reliable judgments and greater transparency for evaluating both human raters and their judgments. Further benefits include reduced need for expert gold, the opportunity for dual-supervision from ratings and rationales, and added value from the rationales themselves.*

- ▷ Paper: Tyler McDonnell, Mucahid Kutlu, Tamer Elsayed, Matthew Lease: The Many Benefits of Annotator Rationales for Relevance Judgments. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Best Sister Conferences. Pages 4909-4913

### [Download paper](#)

*Abstract: When collecting subjective human ratings of items, it can be difficult to measure and enforce data quality due to task subjectivity and lack of insight into how judges arrive at each rating decision. To address this, we propose requiring judges to provide a specific type of rationale underlying each rating decision. We evaluate this approach in the domain of Information Retrieval, where human judges rate the relevance of Webpages. Cost-benefit analysis over 10,000 judgments collected on Mechanical Turk suggests a win-win: experienced crowd workers provide rationales with no increase in task completion time while providing further benefits, including more reliable judgments and greater transparency.*

## 5. Activity Recognition: Video

### 5.1. Survey: Human Activity Recognition from Video

□ **Topic not assigned**

- ▷ Paper: L Onofri, P Soda, M Pechenizkiy, G Iannello: A survey on using domain and contextual knowledge for human activity recognition in video streams. *Expert Systems with Applications*, Volume 63, 30 November 2016, Pages 97-111. Elsevier

[Download paper](#)

*Abstract: Human activity recognition has gained an increasing relevance in computer vision and it can be tackled with either non-hierarchical or hierarchical approaches. The former, also known as single-layered approaches, are those that represent and recognize human activities directly from the extracted descriptors, building a model that distinguishes among the activities contained in the training data. The latter represent and recognize human activities in terms of subevents, which are usually recognized by means of single-layered approaches. Alongside of non-hierarchical and hierarchical approaches, we observe that methods incorporating a priori knowledge and context information on the activity are getting growing interest within the community. In this work we refer to this emerging trend in computer vision as knowledge-based human activity recognition with the objective to cover the lack of a summary of these methodologies. More specifically, we survey methods and techniques used in the literature to represent and integrate knowledge and reasoning into the recognition process. We categorize them as statistical approaches, syntactic approaches and description-based approaches. In addition, we further discuss public and private datasets used in this field to promote their use and to enable the interested readers in finding useful resources. This review ends proposing main future research directions in this field.*

### 5.2. Event Recognition in Unconstrained Videos

□ **Topic not assigned**

- ▷ Paper: Jiang, YG., Bhattacharya, S., Chang, SF. et al.: High-level event recognition in unconstrained videos. *Int J Multimed Info Retr* (2013) 2: 73.

[Download paper](#)

*Abstract: The goal of high-level event recognition is to automatically detect complex high-level events in a given video sequence. This is a difficult task especially when videos are captured under unconstrained conditions by non-professionals. Such videos depicting complex events have limited quality control, and therefore, may include severe camera motion, poor lighting, heavy background clutter, and occlusion. However, due to the fast growing popularity of such videos, especially on the Web, solutions to this problem are in high demands and have attracted great interest from researchers. In this paper, we review current technologies for complex event recognition in unconstrained videos. While the existing solutions vary, we identify common key modules and provide detailed descriptions along with some insights for each of them, including extraction and representation of low-level features across different modalities, classification strategies, fusion techniques, etc. Publicly available benchmark datasets, performance metrics, and related research forums are also described. Finally, we discuss promising directions for future research.*

### 5.3. Video Representation for Action Recognition

□ **Topic not assigned**

- ▷ Paper: Wang, H., Oneata, D., Verbeek, J. et al.: A Robust and Efficient Video Representation for Action Recognition *Int J Comput Vis* (2016) 119: 219.

[Download paper](#)

*Abstract: This paper introduces a state-of-the-art video representation and applies it to efficient action recognition and detection. We first propose to improve the popular dense trajectory features by explicit camera motion estimation. More specifically, we extract feature point matches between frames using SURF descriptors and dense optical flow. The matches are used to estimate a homography with RANSAC. To improve the robustness of homography estimation, a human detector is employed to remove outlier matches from the human body as human motion is not constrained by the camera. Trajectories consistent with the homography are considered as due to camera motion, and thus removed. We also use the homography to cancel out camera motion from the optical flow. This results in significant improvement on motion-based HOF and MBH descriptors. We further explore the recent Fisher vector as an alternative feature encoding approach to the standard bag-of-words (BOW) histogram, and consider different ways to include spatial layout information in these encodings. We present a large and varied set of evaluations, considering (i) classification of short basic actions on six datasets, (ii) localization of such actions in feature-length movies, and (iii) large-scale recognition of complex events. We find that our improved trajectory features significantly outperform previous dense trajectories, and that Fisher vectors are superior to BOW encodings for video recognition tasks. In all three tasks, we show substantial improvements over the state-of-the-art results.*



## 5.4. Single Example Human Activity Recognition in Videos

☒ **Presented by: Raghavendran Tata**

▷ Paper: Mehrsan Javan Roshtkhari, Martin D. Levine, Human activity recognition in videos using a single example, In Image and Vision Computing, Volume 31, Issue 11, 2013, Pages 864-876, ISSN 0262-8856

[Download paper](#)

*Abstract: This paper presents a novel approach for action recognition, localization and video matching based on a hierarchical codebook model of local spatio-temporal video volumes. Given a single example of an activity as a query video, the proposed method finds similar videos to the query in a target video dataset. The method is based on the bag of video words (BOV) representation and does not require prior knowledge about actions, background subtraction, motion estimation or tracking. It is also robust to spatial and temporal scale changes, as well as some deformations. The hierarchical algorithm codes a video as a compact set of spatio-temporal volumes, while considering their spatio-temporal compositions in order to account for spatial and temporal contextual information. This hierarchy is achieved by first constructing a codebook of spatio-temporal video volumes. Then a large contextual volume containing many spatio-temporal volumes (ensemble of volumes) is considered. These ensembles are used to construct a probabilistic model of video volumes and their spatio-temporal compositions. The algorithm was applied to three available video datasets for action recognition with different complexities (KTH, Weizmann, and MSR II) and the results were superior to other approaches, especially in the case of a single training example and cross-dataset. We use this term to denote a query that is selected from a particular dataset when the target videos originate from another dataset. In this situation, the two datasets have been recorded under different conditions. action recognition.*

## 6. Activity Recognition: Emotions & Human-Robot Interaction

### 6.1. Context in Emotion Perception

☒ **Presented by: Niraj Dev Pandey**

▷ Paper: Lisa Feldman Barrett, Batja Mesquita, Maria Gendron: Context in Emotion Perception. Current Directions in Psychological Science Vol 20, Issue 5, pp. 286 - 290 (2011)

[Download paper](#)

*Abstract: We review recent work demonstrating consistent context effects during emotion perception. Visual scenes, voices, bodies, other faces, cultural orientation, and even words shape how emotion is perceived in a face, calling into question the still-common assumption that the emotional state of a person is written on and can be read from the face like words on a page. Incorporating context during emotion perception appears to be routine, efficient, and, to some degree, automatic. This evidence challenges the standard view of emotion perception represented in psychology texts, in the cognitive neuroscience literature, and in the popular media and points to a necessary change in the basic paradigm used in the scientific study of emotion perception.*

▷ Paper: Lisa Feldman Barrett, Elizabeth A. Kensinger: Context Is Routinely Encoded During Emotion Perception. Psychological Science Vol 21, Issue 4, pp. 595 - 599 (2010)

[Download paper](#)

*Abstract: In the present study, we investigated whether context is routinely encoded during emotion perception. For the first time, we show that people remember the context more often when asked to label an emotion in a facial expression than when asked to judge the expression's simple affective significance (which can be done on the basis of the structural features of the face alone). Our findings are consistent with an emerging literature showing that facial muscle actions (i.e., structural features of the face), when viewed in isolation, might be insufficient for perceiving emotion.*

▷ Paper: Katie Hoemann, Maria Gendron, Lisa Feldman Barrett, Mixed emotions in the predictive brain, Current Opinion in Behavioral Sciences, Volume 15, June 2017, Pages 51-57, ISSN 2352-1546

[Download paper](#)

*Abstract: Understanding complex or mixed emotions first requires an exploration of the human nervous system underlying emotions, and indeed all experience. We review current research in neuroscience, which describes the brain as a predictive, internal model of the world that flexibly combines features from past experience to construct emotions. We argue that 'mixed emotions' result when these features of past experience correspond to multiple emotion categories. Integrating event perception and cognitive linguistic theories, we propose that 'mixed emotions' are perceived as an episode of distinct, linked emotional events due to attentional shifts which update the predicted model of experience. These proposed mechanisms have profound implications for the study of emotion; we conclude by suggesting methodological improvements for future research.*

## 6.2. Recognition of Facial Actions

☒ **Presented by: Valerie Chikukwa**

▷ Paper: Bartlett, M. S., Littlewort, G., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006). Automatic recognition of facial actions in spontaneous expressions. *Journal of multimedia*, 1(6), 22-35.

[Download paper](#)

*Abstract: Spontaneous facial expressions differ from posed expressions in both which muscles are moved, and in the dynamics of the movement. Advances in the field of automatic facial expression measurement will require development and assessment on spontaneous behavior. Here we present preliminary results on a task of facial action detection in spontaneous facial expressions. We employ a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). The system automatically detects frontal faces in the video stream and coded each frame with respect to 20 Action units. The approach applies machine learning methods such as support vector machines and AdaBoost, to texture-based image representations. The output margin for the learned classifiers predicts action unit intensity. Frame-by-frame intensity measurements will enable investigations into facial expression dynamics which were previously intractable by human coding.*

## 6.3. Stress Detection using Facial Cues

☒ **Presented by: Mofassir Ul Islam Arif**

▷ Paper: G. Giannakakis, M. Padiaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P.G. Simos, K. Marias, M. Tsiknakis, Stress and anxiety detection using facial cues from videos, In *Biomedical Signal Processing and Control*, Volume 31, 2017, Pages 89-101, ISSN 1746-8094

[Download paper](#)

*Abstract: This study develops a framework for the detection and analysis of stress/anxiety emotional states through video-recorded facial cues. A thorough experimental protocol was established to induce systematic variability in affective states (neutral, relaxed and stressed/anxious) through a variety of external and internal stressors. The analysis was focused mainly on non-voluntary and semi-voluntary facial cues in order to estimate the emotion representation more objectively. Features under investigation included eye-related events, mouth activity, head motion parameters and heart rate estimated through camera-based photoplethysmography. A feature selection procedure was employed to select the most robust features followed by classification schemes discriminating between stress/anxiety and neutral states with reference to a relaxed state in each experimental phase. In addition, a ranking transformation was proposed utilizing self reports in order to investigate the correlation of facial parameters with a participant perceived amount of stress/anxiety. The results indicated that, specific facial cues, derived from eye activity, mouth activity, head movements and camera based heart activity achieve good accuracy and are suitable as discriminative indicators of stress and anxiety.*

## 6.4. Environments & Activity Recognition for Autonomous Robots

☒ **Presented by: Malek Boukhari**

▷ Paper: Muhannad Alomari, Paul Duckworth, Nils Bore, Majd Hawasly, David C. Hogg, Anthony G. Cohn: Grounding of Human Environments and Activities for Autonomous Robots. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Main track. Pages 1395-1402.

[Download paper](#)

*Abstract: With the recent proliferation of human-oriented robotic applications in domestic and industrial scenarios, it is vital for robots to continually learn about their environments and about the humans they share their environments with. In this paper, we present a novel, online, incremental framework for unsupervised symbol grounding in real-world, human environments for autonomous robots. We demonstrate the flexibility of the framework by learning about colours, people names, usable objects and simple human activities, integrating state-of-the-art object segmentation, pose estimation, activity analysis along with a number of sensory input encodings into a continual learning framework. Natural language is grounded to the learned concepts, enabling the robot to communicate in a human-understandable way. We show, using a challenging real-world dataset of human activities as perceived by a mobile robot, that our framework is able to extract useful concepts, ground natural language descriptions to them, and, as a proof-of-concept, generate simple sentences from templates to describe people and the activities they are engaged in.*

▷ Paper: Gori, I., Aggarwal, J. K., Matthies, L., & Ryoo, M. S. (2016). Multitype activity recognition in robot-centric scenarios. *IEEE Robotics and Automation Letters*, 1(1), 593-600.

[Download paper](#)

*Abstract: Activity recognition is very useful in scenarios where robots interact with, monitor, or assist humans. In the past years many types of activities—single actions, two persons interactions or ego-centric activities, to name a few—have been analyzed. Whereas traditional methods treat such types*

of activities separately, an autonomous robot should be able to detect and recognize multiple types of activities to effectively fulfill its tasks. We propose a method that is intrinsically able to detect and recognize activities of different types that happen in sequence or concurrently. We present a new unified descriptor, called relation history image (RHI), which can be extracted from all the activity types we are interested in. We then formulate an optimization procedure to detect and recognize activities of different types. We apply our approach to a new dataset recorded from a robot-centric perspective and systematically evaluate its quality compared to multiple baselines. Finally, we show the efficacy of the RHI descriptor on publicly available datasets performing extensive comparisons.

## 7. Multimodality

### 7.1. Multimodal Interaction

□ **Topic not assigned**

▷ Paper: M Turk: Multimodal interaction: A review. Pattern Recognition Letters, Volume 36, January, 2014, Pages 189-195. Elsevier

[Download paper](#)

Abstract: *People naturally interact with the world multimodally, through both parallel and sequential use of multiple perceptual modalities. Multimodal human-computer interaction has sought for decades to endow computers with similar capabilities, in order to provide more natural, powerful, and compelling interactive experiences. With the rapid advance in non-desktop computing generated by powerful mobile devices and affordable sensors in recent years, multimodal research that leverages speech, touch, vision, and gesture is on the rise. This paper provides a brief and personal review of some of the key aspects and issues in multimodal interaction, touching on the history, opportunities, and challenges of the area, especially in the area of multimodal integration. We review the question of early vs. late integration and find inspiration in recent evidence in biological sensory integration. Finally, we list challenges that lie ahead for research in multimodal human-computer interaction.*

▷ Paper: MR Morris: Web on the wall: insights from a multimodal interaction elicitation study. Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces, Pages 95-104, 2012. ACM

[Download paper](#)

Abstract: *New sensing technologies like Microsoft's Kinect provide a low-cost way to add interactivity to large display surfaces, such as TVs. In this paper, we interview 25 participants to learn about scenarios in which they would like to use a web browser on their living room TV. We then conduct an interaction-elicitation study in which users suggested speech and gesture interactions for fifteen common web browser functions. We present the most popular suggested interactions, and supplement these findings with observational analyses of common gesture and speech conventions adopted by our participants. We also reflect on the design of multimodal, multi-user interaction-elicitation studies, and introduce new metrics for interpreting user-elicitation study findings.*

### 7.2. Gesture & Speech

□ **Topic not assigned**

▷ Paper: P Wagner, Z Malisz, S Kopp: Gesture and speech in interaction: An overview. In: Journal Speech Communication, Volume 57, February, 2014. Pages 209-232. Elsevier

[Download paper](#)

Abstract: *Gestures and speech interact. They are linked in language production and perception, with their interaction contributing to felicitous communication. The multifaceted nature of these interactions has attracted considerable attention from the speech and gesture community. This article provides an overview of our current understanding of manual and head gesture form and function, of the principle functional interactions between gesture and speech aiding communication, transporting meaning and producing speech. Furthermore, we present an overview of research on temporal speech-gesture synchrony, including the special role of prosody in speech-gesture alignment. In addition, we provide a summary of tools and data available for gesture analysis, and describe speech-gesture interaction models and simulations in technical systems. This overview also serves as an introduction to a Special Issue covering a wide range of articles on these topics. We provide links to the Special Issue throughout this paper.*

### 7.3. Multimedia Summarization

□ **Topic not assigned**

▷ Paper: F Metze, D Ding, E Younessian, A Hauptmann: Beyond audio and video retrieval: topic-oriented multimedia summarization. In: International Journal of Multimedia Information (2013) Vol. 2, pages 131-144. Springer

[Download paper](#)

Abstract: *Given the deluge of multimedia content that is becoming available over the Internet, it is*

increasingly important to be able to effectively examine and organize these large stores of information in ways that go beyond browsing or collaborative filtering. In this paper we review previous work on audio and video processing, and define the task of Topic-Oriented Multimedia Summarization (TOMS) using natural language generation: given a set of automatically extracted features from a video (such as visual concepts and ASR transcripts) a TOMS system will automatically generate a paragraph of natural language (“a recounting”), which summarizes the important information in a video belonging to a certain topic area, and provides explanations for why a video was matched and retrieved. We see this as a first step towards systems that will be able to discriminate visually similar, but semantically different videos, compare two videos and provide textual output or summarize a large number of videos at once. In this paper, we introduce our approach of solving the TOMS problem. We extract visual concept features and ASR transcription features from a given video, and develop a template-based natural language generation system to produce a textual recounting based on the extracted features. We also propose possible experimental designs for continuously evaluating and improving TOMS systems, and present results of a pilot evaluation of our initial system.

#### 7.4. Multimodal Ensembles

□ **Topic not assigned**

- ▷ Paper: Maskey, S., & Hirschberg, J. (2005). Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization. In Ninth European Conference on Speech Communication and Technology.

[Download paper](#)

Abstract: We present results of an empirical study of the usefulness of different types of features in selecting extractive summaries of news broadcasts for our Broadcast News Summarization System. We evaluate lexical, prosodic, structural and discourse features as predictors of those news segments which should be included in a summary. We show that a summarization system that uses a combination of these feature sets produces the most accurate summaries, and that a combination of acoustic/prosodic and structural features are enough to build a “good” summarizer when speech transcription is not available.

- ▷ Paper: Sameer Maskey and Julia Hirschberg. 2006. Summarizing speech without text using hidden Markov models. In Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers (NAACL-Short ‘06). Association for Computational Linguistics, Stroudsburg, PA, USA, 89-92.

[Download paper](#)

Abstract: We present a method for summarizing speech documents without using any type of transcript/text in a Hidden Markov Model framework. The hidden variables or states in the model represent whether a sentence is to be included in a summary or not, and the acoustic/prosodic features are the observation vectors. The model predicts the optimal sequence of segments that best summarize the document. We evaluate our method by comparing the predicted summary with one generated by a human summarizer. Our results indicate that we can generate ‘good’ summaries even when using only acoustic/prosodic information, which points toward the possibility of text-independent summarization for spoken documents.

- ▷ Paper: Levitan, S. I., An, G., Ma, M., Levitan, R., Rosenberg, A., & Hirschberg, J. (2016). Combining Acoustic-Prosodic, Lexical, and Phonotactic Features for Automatic Deception Detection. In INTERSPEECH (pp. 2006-2010).

[Download paper](#)

Abstract: Improving methods of automatic deception detection is an important goal of many researchers from a variety of disciplines, including psychology, computational linguistics, and criminology. We present a system to automatically identify deceptive utterances using acoustic-prosodic, lexical, syntactic, and phonotactic features. We train and test our system on the Interspeech 2016 ComParE challenge corpus, and find that our combined features result in performance well above the challenge baseline on the development data. We also perform feature ranking experiments to evaluate the usefulness of each of our feature sets. Finally, we conduct a cross-corpus evaluation by training on another deception corpus and testing on the ComParE corpus.

- ▷ Paper: Mendels, G., Levitan, S. I., Lee, K. Z., & Hirschberg, J. (2017). Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection. Proc. Interspeech 2017, 1472-1476.

[Download paper](#)

Abstract: Automatic deception detection is an important problem with far-reaching implications for many disciplines. We present a series of experiments aimed at automatically detecting deception from speech. We use the Columbia X-Cultural Deception (CXD) Corpus, a large-scale corpus of within-subject deceptive and non-deceptive speech, for training and evaluating our models. We compare the use of spectral, acoustic-prosodic, and lexical feature sets, using different machine learning models. Finally, we design a single hybrid deep model with both acoustic and lexical features trained jointly that achieves

state-of-the-art results on the CXD corpus.

## 7.5. Visual Lecture Transcripts

### ☒ Presented by: reserved

- ▷ Paper: V Shin, F Berthouzoz, W Li, F Durand: Visual transcripts: lecture notes from blackboard-style lecture videos. In: ACM Transactions on Graphics (TOG), Volume 34 Issue 6, November 2015, pages 1-10. ACM

[Download paper](#)

Abstract: *Blackboard-style lecture videos are popular, but learning using existing video player interfaces can be challenging. Viewers cannot consume the lecture material at their own pace, and the content is also difficult to search or skim. For these reasons, some people prefer lecture notes to videos. To address these limitations, we present Visual Transcripts, a readable representation of lecture videos that combines visual information with transcript text. To generate a Visual Transcript, we first segment the visual content of a lecture into discrete visual entities that correspond to equations, figures, or lines of text. Then, we analyze the temporal correspondence between the transcript and visuals to determine how sentences relate to visual entities. Finally, we arrange the text and visuals in a linear layout based on these relationships. We compare our result with a standard video player, and a state-of-the-art interface designed specifically for blackboard-style lecture videos. User evaluation suggests that users prefer our interface for learning and that our interface is effective in helping them browse or search through lecture videos.*

- ▷ Paper: Monserrat, T. J. K. P., Zhao, S., McGee, K., & Pandey, A. V. (2013, April). NoteVideo: facilitating navigation of blackboard-style lecture videos. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1139-1148). ACM.

[Download paper](#)

Abstract: *Khan Academy's pre-recorded blackboard-style lecture videos attract millions of online users every month. However, current video navigation tools do not adequately support the kinds of goals that students typically have, like quickly finding a particular concept in a blackboard-style lecture video. This paper reports on the development and evaluation of the new NoteVideo and its improved version, NoteVideo+, systems for identifying the conceptual 'objects' of a blackboard-based video – and then creating a summarized image of the video and using it as an in-scene navigation interface that allows users to directly jump to the video frame where that object first appeared instead of navigating it linearly through time. The research consisted of iteratively implementing the system and then having users perform four different navigation tasks using three different interfaces: Scrubbing, Transcript, and NoteVideo. Results of the study show that participants perform significantly better on all four tasks while using the NoteVideo and its improved version - NoteVideo+ - as compared to others*

## 7.6. Enhancing Educational Videos

### □ Topic not assigned

- ▷ Paper: Pavel, A., Reed, C., Hartmann, B., & Agrawala, M. (2014, October). Video digests: a browsable, skimmable format for informational lecture videos. In UIST (pp. 573-582).

[Download paper](#)

Abstract: *Increasingly, authors are publishing long informational talks, lectures, and distance-learning videos online. However, it is difficult to browse and skim the content of such videos using current timeline-based video players. Video digests are a new format for informational videos that afford browsing and skimming by segmenting videos into a chapter/section structure and providing short text summaries and thumbnails for each section. Viewers can navigate by reading the summaries and clicking on sections to access the corresponding point in the video. We present a set of tools to help authors create such digests using transcript-based interactions. With our tools, authors can manually create a video digest from scratch, or they can automatically generate a digest by applying a combination of algorithmic and crowdsourcing techniques and then manually refine it as needed. Feedback from first-time users suggests that our transcript-based authoring tools and automated techniques greatly facilitate video digest creation. In an evaluative crowdsourced study we find that given a short viewing time, video digests support browsing and skimming better than timeline-based or transcript-based video players.*

- ▷ Paper: Kim, J., Guo, P. J., Cai, C. J., Li, S. W. D., Gajos, K. Z., & Miller, R. C. (2014, October). Data-driven interaction techniques for improving navigation of educational videos. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 563-572). ACM.

[Download paper](#)

Abstract: *With an unprecedented scale of learners watching educational videos on online platforms such as MOOCs and YouTube, there is an opportunity to incorporate data generated from their interactions into the design of novel video interaction techniques. Interaction data has the potential to help not only instructors to improve their videos, but also to enrich the learning experience of educational video*

watchers. This paper explores the design space of data-driven interaction techniques for educational video navigation. We introduce a set of techniques that augment existing video interface widgets, including: a 2D video timeline with an embedded visualization of collective navigation traces; dynamic and non-linear timeline scrubbing; data-enhanced transcript search and keyword summary; automatic display of relevant still frames next to the video; and a visual summary representing points with high learner activity. To evaluate the feasibility of the techniques, we ran a laboratory user study with simulated learning tasks. Participants rated watching lecture videos with interaction data to be efficient and useful in completing the tasks. However, no significant differences were found in task performance, suggesting that interaction data may not always align with moment-by-moment information needs during the tasks

## 8. Historic Milestones

### 8.1. Enhancing Human Problem Solving

☒ **Presented by: Marvin Forstreuter**

▷ Paper: JCR Licklider: "Man-Computer Symbiosis". Reprinted from IRE (now IEEE) Transactions on Human Factors in Electronics, volume HFE-1, pages 4–11, March 1960.

[Download paper](#)

Abstract: *Man-computer symbiosis is an expected development in cooperative interaction between men and electronic computers. It will involve very close coupling between the human and the electronic members of the partnership. The main aims are 1) to let computers facilitate formulative thinking as they now facilitate the solution of formulated problems, and 2) to enable men and computers to cooperate in making decisions and controlling complex situations without inflexible dependence on predetermined programs. In the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking. Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them. Prerequisites for the achievement of the effective, cooperative association include developments in computer time sharing, in memory components, in memory organization, in programming languages, and in input and output equipment.*

▷ Paper: JCR Licklider: "The Computer as a Communication Device". Reprinted from Science and Technology, April 1968.

Abstract: ...

### 8.2. Sword of Damocles

□ **Topic not assigned**

▷ Paper: Sutherland, Ivan E. A head-mounted three dimensional display. Proceedings of the 1968 Fall Joint Computer Conference, San Francisco, CA, December 1968, Pages 757-764. ACM

[Download paper](#)

Abstract: *The fundamental idea behind the three-dimensional display is to present the user with a perspective image which changes as he moves. The retinal image of the real objects which we see is, after all, only two-dimensional. Thus if we can place suitable two-dimensional images on the observer's retinas, we can create the illusion that he is seeing a three-dimensional object. Although stereo presentation is important to the three-dimensional illusion, it is less important than the change that takes place in the image when the observer moves his head. The image presented by the three-dimensional display must change in exactly the way that the image of a real object would change for similar motions of the user's head. Psychologists have long known that moving perspective images appear strikingly three-dimensional even without stereo presentation; the three-dimensional display described in this paper depends heavily on this "kinetic depth effect."*

### 8.3. Augmenting Human Intellect

□ **Topic not assigned**

▷ Paper: C Engelbart, and WK English: A research center for augmenting human intellect. Proceedings of the 1968 Fall Joint Computer Conference, San Francisco, CA, December 1968, Vol. 33, pp. 395-410. ACM

[Download paper](#)

Abstract: *This paper describes a multisponsor research center at Stanford Research Institute in man-computer interaction.*

### 8.4. Personal Dynamic Media

☒ **Presented by: Kamila Serwa**

▷ Paper: Kay, Alan C. and Adele Goldberg. "Personal Dynamic Media." Computer 10 (1977): 31-41.

[Download paper](#)

Abstract: *The Learning Research Group at Xerox Palo Alto Research Center is concerned with all aspects of the communication and manipulation of knowledge. We design, build, and use dynamic media which can be used by human beings of all ages. Several years ago, we crystallized our dreams into a design idea for a personal dynamic medium the size of a notebook (the Dynabook) which could be owned by everyone and could have the power to handle virtually all of its owner's information-related needs. Towards this goal we have designed and built a communications system: the Smalltalk language, implemented on small computers we refer to as "interim Dynabooks." We are exploring the use of this system as a programming and problem solving tool; as an interactive memory for the storage and manipulation of data; as a text editor; and as a medium for expression through drawing, painting, animating pictures, and composing and generating music.*

#### 8.5. Put That There

□ **Topic not assigned**

- ▷ Paper: RA Bolt: "Put-that-there": Voice and gesture at the graphics interface. SIGGRAPH '80 Proceedings of the 7th annual conference on Computer graphics and interactive techniques, Pages 262-270, 1980.

[Download paper](#)

Abstract: *Recent technological advances in connected-speech recognition and position sensing in space have encouraged the notion that voice and gesture inputs at the graphics interface can converge to provide a concerted, natural user modality. The work described herein involves the user commanding simple shapes about a large-screen graphics display surface. Because voice can be augmented with simultaneous pointing, the free usage of pronouns becomes possible, with a corresponding gain in naturalness and economy of expression. Conversely, gesture aided by voice gains precision in its power to reference.*

#### 8.6. Tangible Bits

□ **Topic not assigned**

- ▷ Paper: Ishii, Hiroshi, and Brygg Ullmer. "Tangible bits: towards seamless interfaces between people, bits and atoms." Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. ACM, 1997.

[Download paper](#)

Abstract: *This paper presents our vision of Human Computer Interaction (HCI): "Tangible Bits." Tangible Bits allows users to "grasp & manipulate" bits in the center of users' attention by coupling the bits with everyday physical objects and architectural surfaces. Tangible Bits also enables users to be aware of background bits at the periphery of human perception using ambient display media such as light, sound, airflow, and water movement in an augmented space. The goal of Tangible Bits is to bridge the gaps between both cyberspace and the physical environment, as well as the foreground and background of human activities. This paper describes three key concepts of Tangible Bits: interactive surfaces; the coupling of bits with graspable physical objects; and ambient media for background awareness. We illustrate these concepts with three prototype systems – the metaDESK, transBOARD and ambientROOM – to identify underlying research issues.*

### 9. Transdisciplinary Issues

#### 9.1. Systemic-Functional Theory of Language: Semantic Networks

□ **Topic not assigned**

- ▷ Paper: R Hasan, C Cloran, G Williams and A Lukin: Semantic networks: the description of linguistic meaning in SFL. In R. Hasan, C. MIM. Matthiessen & J. J. Webster (eds), Continuing Discourse on Language: A Functional Perspective Volume 2, 1 edn, pp. 697-738, Equinox Publishing Ltd, London, UK and Oakville, USA, 2005.

[Download paper](#)

Abstract: *In SFL, the semantics of a language calls for as much attention as its lexicogrammar: in fact, meaning and wording are two sides of the same coin; the description of both is equally central to understanding 'how language works' (Halliday, McIntosh and Strevens, 1964) – which has been Halliday's agenda since the beginning of his engagement with linguistics. But what actually led SFL into the exploration of semantics as a legitimate domain for description was not these theoretical considerations, per se; rather, like other aspects of the evolution of SFL, interest in semantics too arose in attempts to resolve certain problems in the course of research during the 1960s. This chapter presents one perspective on the course of this development, specifically with respect to semantic networks as a resource for the analysis of meaning.*

#### 9.2. Systemic-Functional Theory of Language: Academic Writing

□ **Topic not assigned**

- ▷ Paper: D Motta-Roth: The Role of Context in Academic Text Production and Writing Pedagogy. In: C Bazerman, A Bonini, D Figueiredo (eds), Genre in a Changing World, pp. 317-336, Parlor

Press, 2009.

[Download paper](#)

*Abstract: The problem of text production in academic genres has been a challenge for undergraduate and graduate students as well as for writing teachers from different departments. Previous research has provided important results on the structural aspects of academic genres (Swales, 1990) and the discursive construction of identity in academic writing (Ivanic, 1998). However, few studies have concentrated on the contributions of exploration and reflection on context to actual teaching practices. From the perspective of Systemic-Functional Linguistics (SFL), in this paper I would like to focus on the reciprocal relationship between text and context, i.e., the way context can be recreated by analysis of text and vice versa. The aim is to point out some practical implications derived from the use of SFL principles in academic writing teaching and research through context exploration. The focus will be on writing activities that aim at fostering students' awareness about the connections between contextual features (activity, identity, relations as well as the role performed by text in the situation) and their respective linguistic realizations (expression of content, instantiation of relationships between interlocutors, and organization of text). One of the main challenges in language education and research is to teach creative ways to negotiate the norms of the language system (grammar) within the academic culture: the set of meanings, rules, values, power relations and relevant genres that constitute the social practices of a community. Educating students about the uses of language in specific contexts depends on clear descriptions of the connections between text and context. Public discourse on academic publication in Brazil is mainly issued by the Ministry of Education through its two main Research Funding Agencies, CAPES and CNPq, which hold quantitative and qualitative expectations about scholars' intellectual production but offer no substantial line of financial support for pedagogic research and course development. Thus full-fledged writing programs are seldom found in Brazilian universities. Very often what we find is some individual or collective teaching initiatives that have survived defying all the odds (e.g., lack of personnel and financial resources) situated in specific institutions.*

### 9.3. Activity Theory: Taxonomy of Use in HCI

☒ **Presented by: Lucky Okehigbemen**

▷ Paper: T Clemmensen, V Kaptelinin, B Nardi: Making HCI theory work: an analysis of the use of activity theory in HCI research – Behaviour & Information Technology Vol. 35 , Iss. 8, pp. 608-627, 2016. Taylor & Francis

[Download paper](#)

*Abstract: This paper reports a study of the use of activity theory in human-computer interaction (HCI) research. We analyse activity theory in HCI since its first appearance about 25 years ago. Through an analysis and meta-synthesis of 109 selected HCI activity theory papers, we created a taxonomy of 5 different ways of using activity theory: (1) analysing unique features, principles, and problematic aspects of the theory; (2) identifying domain-specific requirements for new theoretical tools; (3) developing new conceptual accounts of issues in the field of HCI; (4) guiding and supporting empirical analyses of HCI phenomena; and (5) providing new design illustrations, claims, and guidelines. We conclude that HCI researchers are not only users of imported theory, but also theory-makers who adapt and develop theory for different purposes.*

### 9.4. Activity Theory: Plans as Situated Action

☐ **Topic not assigned**

▷ Paper: Bardram, Jakob E. "Plans as situated action: an activity theory approach to workflow systems." Proceedings of the Fifth European Conference on Computer Supported Cooperative Work. Springer Netherlands, 1997.

[Download paper](#)

*Abstract: Within the community of CSCW the notion and nature of workflow systems as prescriptions of human work has been debated and criticised. Based on the work of Suchman (1987) the notion of situated action has often been viewed as opposed to planning work. Plans, however, do play an essential role in realising work. Based on experiences from designing a computer system that supports the collaboration within a hospital, this paper discusses how plans themselves are made out of situated action, and in return are realised in situ. Thus, work can be characterised as situated planning. This understanding is backed up by Activity Theory, which emphasises the connection between plans and the contextual conditions for realising these plans in actual work.*

## Source

Texts can be downloaded from the [Learnweb-course](#).

List of the 21 topics which have not been assigned yet: 2.4., 3.1., 3.4., 4.2., 4.3., 4.4., 5.1., 5.2., 5.3., 7.1., 7.2., 7.3., 7.4., 7.6., 8.2., 8.3., 8.5., 8.6., 9.1., 9.2., 9.4., Please see individual papers for a description.