1 Introduction

linearize: {collections, technologies, policies} X {where we are, what we need, why important}
collections: collections, archiving, standards
technologies
policies: copyright, IPR, privacy, moral rights

2 Description of the issues

Mark, Doug, Marcello

2.1 Types of spoken word collection

2.2 User communities

2.3 Archives

1. analog / historical
2. digital / future

2.4 Technology - review of state-of-the-art

2.5 Access

1. Current - physical, some web
2. Future - perhaps physical, web and other technologies

2.6 Magnitude

- current estimates
- anticipated growth

2.7 Opportunities

- Convergence of need
- maturity of technology

3 Technologies and new landscapes

Marcello, Doug, Lori, Franciska, Fabrizio, [Clare]

Introduction links to research approaches (data driven)
3.1 audio/signal processing
3.2 speech and speaker recognition
3.3 information extraction (eg NEs)
3.4 Language ID
3.5 collection level browsing and search
3.6 Presentation (eg Summarization)

4 Privacy and copyright

Clare, Richard, [Doug, Franciska]

4.1 Copyright
1. Impact variews by collection
2. Poss exemption for educational institutions

4.2 Privacy (900lb gorilla, acknowledge it, can’t ignore, can’t solve)

4.3 National security
1. Expectation of privacy (esp conversations)

5 Problems associated with archiving and access (collecting and managing)

Carl, Richard, [Lori, Franciska, Steve B]

5.1 Preservation
1. Standards for preservation
   (a) Digital data preservation
   (b) File formats
2. Sustainable digital repositories
   (a) funding
   (b) migration
   (c) refreshing/augmenting
3. Capture and preservation of collections, web harvesting

5.2 Content structure
1. Administrative and structural metadata
2. Descriptive metadata (transcripts, ...)
3. Metadata representations and standards

6 Recommendations

Jerry, Steve

A Glossary
B Use cases
C Technical details
D Additional sources