Spoken Content Retrieval: Challenges, Techniques and Applications

(Part 1: Introduction and Overview)

Gareth J. F. Jones

Centre for Next Generation Localisation
School of Computing, Dublin City University, Dublin, Ireland

Overview

sfi Fondúireacht Eolaíochta Éireann Science Foundation Ireland
DCU
CNGL Centre for Next Generation Localisation

Introduction to Spoken Content Retrieval (SCR)

- Spoken Content Retrieval (SCR) provides users with access to digitized audio-visual content with a spoken language component.

Introduction to Spoken Content Retrieval (SCR)

- Spoken Content Retrieval (SCR) provides users with access to digitized audio-visual content with a spoken language component.

This course is based on:

- Martha Larson and Gareth J. F. Jones (2012)
  Spoken content retrieval: A survey of techniques and technologies
  Foundations and Trends in Information Retrieval, 5 (4-5). pp. 235-422.
  Available from: `http://doras.dcu.ie/17158/`.

Introduction to Spoken Content Retrieval (SCR)

- Why is SCR an active research topic in 2013?:

Introduction to Spoken Content Retrieval (SCR)

- Why is SCR an active research topic in 2013?:
  - Unprecedented volume of stored digital spoken content appearing online, and in institutional, enterprise, personal archives.

Introduction to Spoken Content Retrieval (SCR)

- Why is SCR an active research topic in 2013?:
  - Unprecedented volume of stored digital spoken content appearing online, and in institutional, enterprise, personal archives.
  - Increasingly diverse sources, e,g, spoken and audio-visual content, but also other settings, e.g. audio annotations of images on a smartphone.
    Growth in content which is spontaneous or conversational and/or recorded outside a studio setting.

Introduction to Spoken Content Retrieval (SCR)

- Why is SCR an active research topic in 2013?:
  - Unprecedented volume of stored digital spoken content appearing online, and in institutional, enterprise, personal archives.
  - Increasingly diverse sources, e,g, spoken and audio-visual content, but also other settings, e.g. audio annotations of images on a smartphone.
    Growth in content which is spontaneous or conversational and/or recorded outside a studio setting.
  - Great variety of needs to search spoken content: factual information, e.g. TV documentaries, lectures; historical records, e.g. meetings, interviews; entertainment, e.g. moves, user-generated internet video.

Introduction to Spoken Content Retrieval (SCR)

- Why is SCR an active research topic in 2013?:
  - Unprecedented volume of stored digital spoken content appearing online, and in institutional, enterprise, personal archives.
  - Increasingly diverse sources, e,g, spoken and audio-visual content, but also other settings, e.g. audio annotations of images on a smartphone.
    Growth in content which is spontaneous or conversational and/or recorded outside a studio setting.
  - Great variety of needs to search spoken content: factual information, e.g. TV documentaries, lectures; historical records, e.g. meetings, interviews; entertainment, e.g. moves, user-generated internet video.
  - Increasing recording of spoken interactions, e.g. call centres. Access to speech media should be intuitive, reliable and comfortable.

Introduction to Spoken Content Retrieval (SCR)

- These factors mean that the present time is one of unprecedented potential for innovative news applications of SCR.

Introduction to Spoken Content Retrieval (SCR)

- These factors mean that the present time is one of unprecedented potential for innovative news applications of SCR.
- Search engines and retrieval systems that make of SCR are better able to connect users with audio-visual content which match their needs for information and content.

Introduction to Spoken Content Retrieval (SCR)

- The basic technologies of SCR are text information retrieval (IR) models and automatic speech recogntiion (ASR).
- ASR systems generate text transcripts from spoken audio.

Introduction to Spoken Content Retrieval (SCR)

- The basic technologies of SCR are text information retrieval (IR) models and automatic speech recogntiion (ASR).
- ASR systems generate text transcripts from spoken audio.
- A naive SCR system would simply use ASR text transcripts to replace text files in a text IR system.

Introduction to Spoken Content Retrieval (SCR)

- The basic technologies of SCR are text information retrieval (IR) models and automatic speech recogntiion (ASR).
- ASR systems generate text transcripts from spoken audio.
- A naive SCR system would simply use ASR text transcripts to replace text files in a text IR system.
- The nature of spoken content, human interaction with spoken content and limitations of ASR systems mean that deeper consideration is needed to realise effect SCR systems.

Definition of Spoken Content Retrieval (SCR)

- **Spoken Content Retrieval (SCR)** is the task of returning speech media results that are relevant to an information need expressed as a user query.

  Includes both audio-only speech content and speech media in other forms, including as part of video.

  - **Speech Retrieval (SR)** - first name used for this technology
  - **Spoken Document Retrieval (SDR)** - later name, but limited in scope - based on defined document units

Definition of Spoken Content Retrieval (SCR)

- **Spoken Content Retrieval (SCR)** is the task of returning speech media results that are relevant to an information need expressed as a user query.

  Includes both audio-only speech content and speech media in other forms, including as part of video.

  - **Speech Retrieval (SR)** - first name used for this technology
  - **Spoken Document Retrieval (SDR)** - later name, but limited in scope - based on defined document units
- **Voice Search** or **Voice Retrieval** focuses on returning textual or speech media based on **spoken** queries.
  - Raises different challenges to SCR, and is outside the scope of this course.

Differences between SCR and text IR

- ASR transcripts generally contain errors. Error rate varies, but can be more than 50%! Such error rates will impact on SCR effectiveness.

Differences between SCR and text IR

- ASR transcripts generally contain errors. Error rate varies, but can be more than 50%! Such error rates will impact on SCR effectiveness.
- Spoken content rarely has the defined structure of text, e.g. paragraphs, sentences.

Differences between SCR and text IR

- ASR transcripts generally contain errors. Error rate varies, but can be more than 50%! Such error rates will impact on SCR effectiveness.
- Spoken content rarely has the defined structure of text, e.g. paragraphs, sentences.
- Spoken content is often informal and not segmented into clearly defined documents, e.g. consider news stories vs a recorded meeting.

Differences between SCR and text IR

- ASR transcripts generally contain errors. Error rate varies, but can be more than 50%! Such error rates will impact on SCR effectiveness.
- Spoken content rarely has the defined structure of text, e.g. paragraphs, sentences.
- Spoken content is often informal and not segmented into clearly defined documents, e.g. consider news stories vs a recorded meeting.
- Speech is a temporal medium, meaning that accessing raw spoken content is time consuming and inefficient. SCR systems must provide visualizations of content to enable easy relevance assessment and to access suitable playback start points.

Differences between SCR and text IR

- ASR technology can generate information not included in standard text media, e.g.

Differences between SCR and text IR

- ASR technology can generate information not included in standard text media, e.g.
  - time code - start/end point of the word within the speech media.

Differences between SCR and text IR

- ASR technology can generate information not included in standard text media, e.g.
  - time code - start/end point of the word within the speech media.
  - statistical confidence that the word has been recognised correctly.

Differences between SCR and text IR

- ASR technology can generate information not included in standard text media, e.g.
  - time code - start/end point of the word within the speech media.
  - statistical confidence that the word has been recognised correctly.
  - ASR system generates multiple hypotheses of the word that may have been spoken, but were less likely than the one included in the transcript.

Differences between SCR and text IR

- ASR technology can generate information not included in standard text media, e.g.
  - time code - start/end point of the word within the speech media.
  - statistical confidence that the word has been recognised correctly.
  - ASR system generates multiple hypotheses of the word that may have been spoken, but were less likely than the one included in the transcript.
  - ASR system can generate **rich transcripts** containing additional information, e.g. speaker change points, male/female speaker, identifying named speaker or audio events, such as applause and laughter.

Relationship of SCR to Speech Recognition

- Generally two distinct branches of speech recognition research:

Relationship of SCR to Speech Recognition

- Generally two distinct branches of speech recognition research:
  - Speech Understanding (SU)): dialogue systems with some notion of “understanding”, e.g. to enquire about train schedules.

Relationship of SCR to Speech Recognition

- Generally two distinct branches of speech recognition research:
  - Speech Understanding (SU)): dialogue systems with some notion of “understanding”, e.g. to enquire about train schedules.
  - Speech Transcription: a “listening typewriter” designed to transcribe exactly what is said in the recorded audio.

Relationship of SCR to Speech Recognition

- Generally two distinct branches of speech recognition research:
  - Speech Understanding (SU)): dialogue systems with some notion of “understanding”, e.g. to enquire about train schedules.
  - Speech Transcription: a “listening typewriter” designed to transcribe exactly what is said in the recorded audio.
- Recent work on:
  - Spoken Term Detection (STD) returns instances of particular words being pronounced within the speech stream;
  - related to Spoken Utterance Retrieval (SUR) returns short speech segments in which specific words are pronounced.

Relationship of SCR to Speech Recognition

- Generally two distinct branches of speech recognition research:
  - Speech Understanding (SU)): dialogue systems with some notion of “understanding”, e.g. to enquire about train schedules.
  - Speech Transcription: a “listening typewriter” designed to transcribe exactly what is said in the recorded audio.
- Recent work on:
  - Spoken Term Detection (STD) returns instances of particular words being pronounced within the speech stream;
  - related to Spoken Utterance Retrieval (SUR) returns short speech segments in which specific words are pronounced.

SCR typically based on speech transcription, but STD and related techniques can be useful.

SCR and Speech Recognition

- ASR transcription systems have a predefined vocabulary - meaning that if a word outside this vocabulary is spoken, it cannot be recognized correctly.

SCR and Speech Recognition

- ASR transcription systems have a predefined vocabulary - meaning that if a word outside this vocabulary is spoken, it cannot be recognized correctly.
  - This can be important for SCR where unusual words can be important to achieve effective retrieval.

SCR and Speech Recognition

- ASR transcription systems have a predefined vocabulary - meaning that if a word outside this vocabulary is spoken, it cannot be recognized correctly.
  - This can be important for SCR where unusual words can be important to achieve effective retrieval.
- Adapting the vocabulary can be difficult, and all content would need to be retranscribed with the new system, which is generally uneconomic or impractical.

SCR and Speech Recognition

- ASR transcription systems have a predefined vocabulary - meaning that if a word outside this vocabulary is spoken, it cannot be recognized correctly.
  - This can be important for SCR where unusual words can be important to achieve effective retrieval.
- Adapting the vocabulary can be difficult, and all content would need to be retranscribed with the new system, which is generally uneconomic or impractical.
- STD technologies are designed to find instances of individual words which can be defined easily for search.

SCR and Speech Recognition

- ASR transcription systems have a predefined vocabulary - meaning that if a word outside this vocabulary is spoken, it cannot be recognized correctly.
  - This can be important for SCR where unusual words can be important to achieve effective retrieval.
- Adapting the vocabulary can be difficult, and all content would need to be retranscribed with the new system, which is generally uneconomic or impractical.
- STD technologies are designed to find instances of individual words which can be defined easily for search.
- Combinations of ASR transcriptions and STD can provide solutions for SCR.

Overview

sfi Fondúireacht Eolaíochta Éireann Science Foundation Ireland
DCU
CNGL Centre for Next Generation Localisation

A Brief Overview of SCR Research

- Research in SCR and underlying technologies ongoing for more than twenty years.

A Brief Overview of SCR Research

- Research in SCR and underlying technologies ongoing for more than twenty years.
- SCR research during this period can be informally divided into a number of phases:
  - Proto-SCR
  - Dawn of SCR
  - Rise of the SCR benchmark
  - Spontaneous, conversational speech

  These phases have involved various technological developments in SCR.

A Brief Overview of SCR Research

- Research in SCR and underlying technologies ongoing for more than twenty years.
- SCR research during this period can be informally divided into a number of phases:
  - Proto-SCR
  - Dawn of SCR
  - Rise of the SCR benchmark
  - Spontaneous, conversational speech

  These phases have involved various technological developments in SCR.
- The technologies explored in this research have been applied in a range of application areas of generally increasing complexity, and investigated for a large number of languages.

History of SCR Research

Proto-SCR

- Mainly in the early 1990s.
- ASR transcription systems not yet available; speech recognition using fixed-vocabulary wordspotting. Recognised small fixed set of words, typically 20-50 carefully chosen words relating to a specific domain.

History of SCR Research

Proto-SCR

- Mainly in the early 1990s.
- ASR transcription systems not yet available; speech recognition using fixed-vocabulary wordspotting. Recognised small fixed set of words, typically 20-50 carefully chosen words relating to a specific domain.
- Used for filtering and classifying spoken items, e.g. voice messages.
- Referred to in the literature as “information retrieval”, but not IR as generally understood, topics are pre-defined, more akin to what is known as “information filtering”.

History of SCR Research

Dawn of SCR

- Roughly 1992-1997
- Standard IR where queries are not specified in advance.

History of SCR Research

Dawn of SCR

- Roughly 1992-1997
- Standard IR where queries are not specified in advance.
- New speech recognition technologies: subword transcription (with uncertainty), subword (phone) lattices - no vocabulary restrictions.

History of SCR Research

Dawn of SCR

- Roughly 1992-1997
- Standard IR where queries are not specified in advance.
- New speech recognition technologies: subword transcription (with uncertainty), subword (phone) lattices - no vocabulary restrictions.
- ASR transcription now available.

History of SCR Research

Dawn of SCR

- Roughly 1992-1997
- Standard IR where queries are not specified in advance.
- New speech recognition technologies: subword transcription (with uncertainty), subword (phone) lattices - no vocabulary restrictions.
- ASR transcription now available.
- Hybrid indexing using ASR transcriptions (accuracy) and phone lattices (vocabulary coverage).

History of SCR Research

Dawn of SCR

- Roughly 1992-1997
- Standard IR where queries are not specified in advance.
- New speech recognition technologies: subword transcription (with uncertainty), subword (phone) lattices - no vocabulary restrictions.
- ASR transcription now available.
- Hybrid indexing using ASR transcriptions (accuracy) and phone lattices (vocabulary coverage).
- Carnegie Mellon Informedia system - first large scale digital library for video search driven by manually-generated closed captions and ASR transcriptions.

History of SCR Research

Dawn of SCR

- Roughly 1992-1997
- Standard IR where queries are not specified in advance.
- New speech recognition technologies: subword transcription (with uncertainty), subword (phone) lattices - no vocabulary restrictions.
- ASR transcription now available.
- Hybrid indexing using ASR transcriptions (accuracy) and phone lattices (vocabulary coverage).
- Carnegie Mellon Informedia system - first large scale digital library for video search driven by manually-generated closed captions and ASR transcriptions.
- Small scale evaluations at individual research sites.

History of SCR Research

Rise of the SCR benchmark

- Text REtrieval Conference (TREC) Spoken Document Retrieval (SDR) task - 1997-2000

History of SCR Research

Rise of the SCR benchmark

- Text REtrieval Conference (TREC) Spoken Document Retrieval (SDR) task - 1997-2000
  - Search of radio and TV news stories.
  - Enabled comparison of retrieval behaviour across sites for the first time.

History of SCR Research

Rise of the SCR benchmark

- Text REtrieval Conference (TREC) Spoken Document Retrieval (SDR) task - 1997-2000
  - Search of radio and TV news stories.
  - Enabled comparison of retrieval behaviour across sites for the first time.
  - Task relatively straightforward - problem declared generally solved in 2000.

History of SCR Research

Rise of the SCR benchmark

- Text REtrieval Conference (TREC) Spoken Document Retrieval (SDR) task - 1997-2000
  - Search of radio and TV news stories.
  - Enabled comparison of retrieval behaviour across sites for the first time.
  - Task relatively straightforward - problem declared generally solved in 2000.
  - Generally based on ASR transcriptions.
    - Extensive use of query expansion.
    - Exploration of document expansion.
    - Investigation of contemporaneous text collections.

History of SCR Research

Spontaneous, conversational speech

- 2005 - present

History of SCR Research

Spontaneous, conversational speech

- 2005 - present
- Searching Spontaneous Conversational Speech workshop series

History of SCR Research

Spontaneous, conversational speech

- 2005 - present
- Searching Spontaneous Conversational Speech workshop series
- CLEF Cross-Language Speech Retrieval 2005-2007
  - Challenging task searching unstructured interviews.
  - Highlighted the need to exploit any available text metadata.

History of SCR Research

Spontaneous, conversational speech

- 2005 - present
- Searching Spontaneous Conversational Speech workshop series
- CLEF Cross-Language Speech Retrieval 2005-2007
  - Challenging task searching unstructured interviews.
  - Highlighted the need to exploit any available text metadata.
- NTCIR SpokenDoc 2011 - present
  - Search of lecture recordings.
  - Searches for relevant segments within lectures.

History of SCR Research

Spontaneous, conversational speech

- 2005 - present
- Searching Spontaneous Conversational Speech workshop series
- CLEF Cross-Language Speech Retrieval 2005-2007
  - Challenging task searching unstructured interviews.
  - Highlighted the need to exploit any available text metadata.
- NTCIR SpokenDoc 2011 - present
  - Search of lecture recordings.
  - Searches for relevant segments within lectures.
- MediaEval speech research tasks 2010 - present
  - Search of semi-professional internet video.
  - Collection developed using crowdsourcing.
  - Focus on identifying jump-in points to begin playback of content.

Use Scenarios

- SCR has been applied in a range of application areas.

Use Scenarios

- SCR has been applied in a range of application areas.
  - Broadcast news retrieval: scripted and careful interviews, “self describing” for a general audience, limited topical scope, clear underlying structure, speech technology easily tuned to the domain.

Use Scenarios

- SCR has been applied in a range of application areas.
  - Broadcast news retrieval: scripted and careful interviews, “self describing” for a general audience, limited topical scope, clear underlying structure, speech technology easily tuned to the domain.
  - Voice and video mail retrieval: frequently informal speaking styles, uncontrolled domain, structure variable, hard to tune speech technology to the task, messages typically short, users may have difficulty recalling suitable search words.

Use Scenarios

- SCR has been applied in a range of application areas.
  - Broadcast news retrieval: scripted and careful interviews, “self describing” for a general audience, limited topical scope, clear underlying structure, speech technology easily tuned to the domain.
  - Voice and video mail retrieval: frequently informal speaking styles, uncontrolled domain, structure variable, hard to tune speech technology to the task, messages typically short, users may have difficulty recalling suitable search words.
  - Search of meetings: informal, challenging speech conditions, may be possible to tune speech technology, complex content structure, what is the appropriate retrieval unit.

Use Scenarios

- SCR has been applied in a range of application areas.
  - Broadcast news retrieval: scripted and careful interviews, “self describing” for a general audience, limited topical scope, clear underlying structure, speech technology easily tuned to the domain.
  - Voice and video mail retrieval: frequently informal speaking styles, uncontrolled domain, structure variable, hard to tune speech technology to the task, messages typically short, users may have difficulty recalling suitable search words.
  - Search of meetings: informal, challenging speech conditions, may be possible to tune speech technology, complex content structure, what is the appropriate retrieval unit.
  - Other examples: call centre recordings, collections of interviews, historical archives, lectures, podcasts, political speeches.

Overview

sfi Fondúireacht Eolaíochta Éireann Science Foundation Ireland
DCU
CNGL Centre for Next Generation Localisation

Architecture of an SCR System

The underlying architecture of an SCR system consists of a more or less stable set of components.

Architecture of an SCR System

The underlying architecture of an SCR system consists of a more or less stable set of components.

- *Query*: User input to the SCR system.

Architecture of an SCR System

The underlying architecture of an SCR system consists of a more or less stable set of components.

- *Query*: User input to the SCR system.
- *Timed Media Index*: Contains features that represent the items in the collections.

Architecture of an SCR System

The underlying architecture of an SCR system consists of a more or less stable set of components.

- *Query*: User input to the SCR system.
- *Timed Media Index*: Contains features that represent the items in the collections.
- *Speech Recognition System*: processes items in the *Speech Collection* at indexing time.

Architecture of an SCR System

The underlying architecture of an SCR system consists of a more or less stable set of components.

- *Query*: User input to the SCR system.
- *Timed Media Index*: Contains features that represent the items in the collections.
- *Speech Recognition System*: processes items in the *Speech Collection* at indexing time.
- *Retrieval System*: matches the query with the items in the collection.

Architecture of an SCR System

- *Ranked Results*: A set of results ordered by their likelihood of potential relevance to the query.

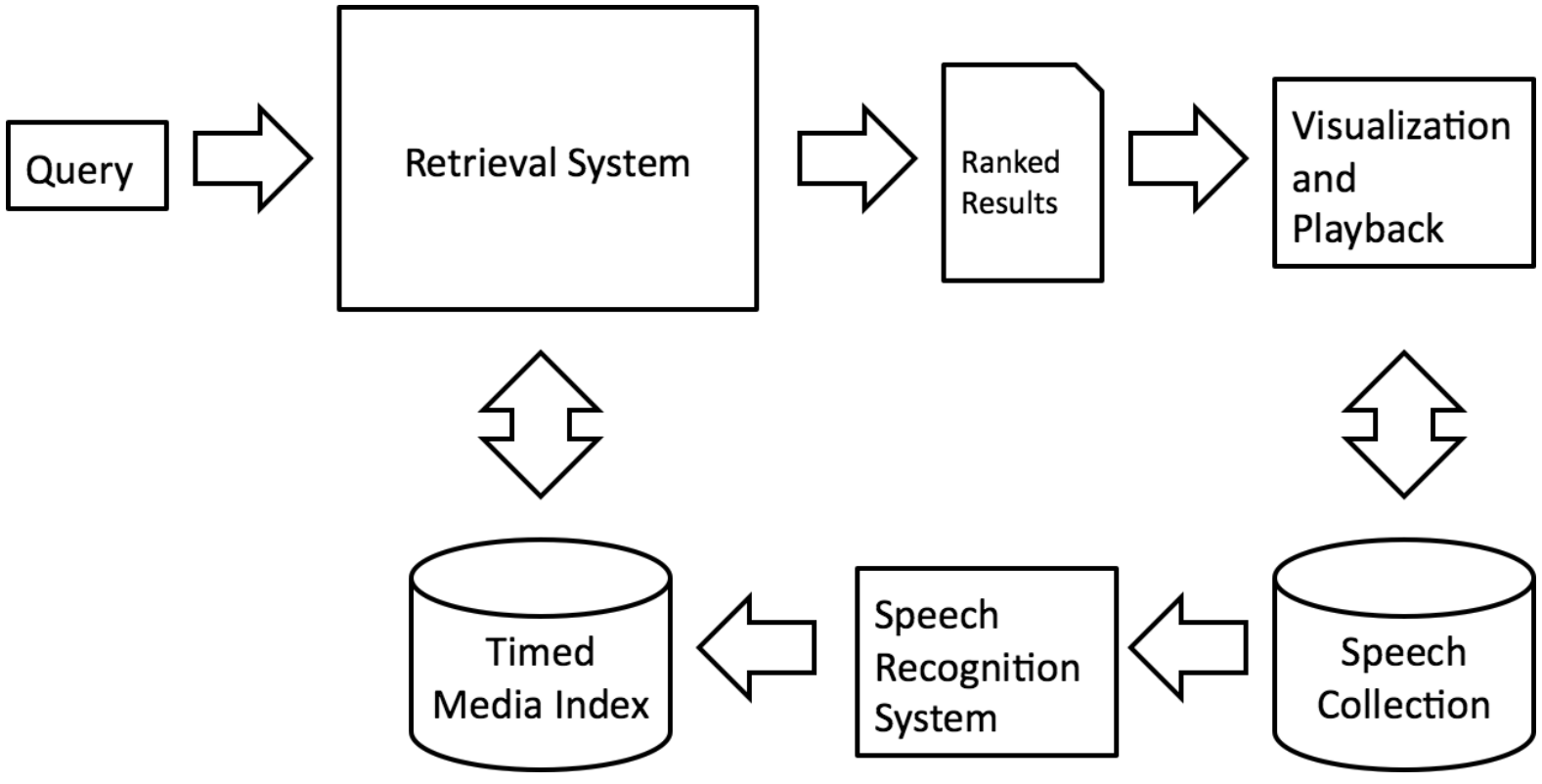
Architecture of an SCR System

- *Ranked Results*: A set of results ordered by their likelihood of potential relevance to the query.
  - Item can be: complete spoken item, fixed segment of speech, segment of speech taking into account the query ("relevance interval"), jump-in time point to start playback.

Architecture of an SCR System

- *Ranked Results*: A set of results ordered by their likelihood of potential relevance to the query.
  - Item can be: complete spoken item, fixed segment of speech, segment of speech taking into account the query ("relevance interval"), jump-in time point to start playback.
- *Visualization and Playback*: enable user to judge relevance without extensive listening, access information efficiently in selected item.

Architecture of an SCR System



Overview

sfi Fondúireacht Eolaíochta Éireann Science Foundation Ireland
DCU
CNGL Centre for Next Generation Localisation

Information Retrieval for SCR

- In very general terms, modern Information Retrieval (IR) systems match user queries against the available content to produce a list ranked by likelihood of relevance.
- The list is sorted based on *ranking score* computed using a *ranking function*.

Information Retrieval for SCR

- In very general terms, modern Information Retrieval (IR) systems match user queries against the available content to produce a list ranked by likelihood of relevance.
- The list is sorted based on *ranking score* computed using a *ranking function*.
- The ranking function can be based on one of a number of IR models including:
  - Vector-space model
  - Probabilistic model
  - Language modelling

Information Retrieval for SCR

- In very general terms, modern Information Retrieval (IR) systems match user queries against the available content to produce a list ranked by likelihood of relevance.
- The list is sorted based on *ranking score* computed using a *ranking function*.
- The ranking function can be based on one of a number of IR models including:
  - Vector-space model
  - Probabilistic model
  - Language modelling
- Objective of any model is to differentiate relevant from non-relevant content (documents).

Information Retrieval for SCR

- In very general terms, modern Information Retrieval (IR) systems match user queries against the available content to produce a list ranked by likelihood of relevance.
- The list is sorted based on *ranking score* computed using a *ranking function*.
- The ranking function can be based on one of a number of IR models including:
  - Vector-space model
  - Probabilistic model
  - Language modelling
- Objective of any model is to differentiate relevant from non-relevant content (documents).
- Typically uses features which are *representative* and *discriminative*.

Relevance Feedback and Query Expansion

- Query expansion via relevance feedback aims to enrich the user's query to better represent their underlying information need using the contents of the search collection.

Relevance Feedback and Query Expansion

- Query expansion via relevance feedback aims to enrich the user's query to better represent their underlying information need using the contents of the search collection.
- Often effective for text IR, but for SCR in addition, can expand query to include words well recognised by the ASR system, which can help to overcome:

Relevance Feedback and Query Expansion

- Query expansion via relevance feedback aims to enrich the user's query to better represent their underlying information need using the contents of the search collection.
- Often effective for text IR, but for SCR in addition, can expand query to include words well recognised by the ASR system, which can help to overcome:
  - problems with query words not well modeled by the ASR system;

Relevance Feedback and Query Expansion

- Query expansion via relevance feedback aims to enrich the user's query to better represent their underlying information need using the contents of the search collection.
- Often effective for text IR, but for SCR in addition, can expand query to include words well recognised by the ASR system, which can help to overcome:
  - problems with query words not well modeled by the ASR system;
  - words outside the vocabulary of the ASR system.

Relevance Feedback and Query Expansion

- Query expansion via relevance feedback aims to enrich the user's query to better represent their underlying information need using the contents of the search collection.
- Often effective for text IR, but for SCR in addition, can expand query to include words well recognised by the ASR system, which can help to overcome:
  - problems with query words not well modeled by the ASR system;
  - words outside the vocabulary of the ASR system.
- Query expansion has proven particularly effective and robust for SCR.

Integrating additional information

- Data fusion: linear combination of ranking scores from multiple sources can be very useful: combines multiple lists created using (independent) content representation, e.g. alternative ASR systems/methods.

Integrating additional information

- Data fusion: linear combination of ranking scores from multiple sources can be very useful: combines multiple lists created using (independent) content representation, e.g. alternative ASR systems/methods.
- “Learning-to-rank” methods can be applied - importance of signals or alternative signals may be relevant - I’m not aware of SCR work that has explored this in practice.

Integrating additional information

- Data fusion: linear combination of ranking scores from multiple sources can be very useful: combines multiple lists created using (independent) content representation, e.g. alternative ASR systems/methods.
- “Learning-to-rank” methods can be applied - importance of signals or alternative signals may be relevant - I’m not aware of SCR work that has explored this in practice.
- Other ranking features developed for text IR can be applied for SCR:

Integrating additional information

- Data fusion: linear combination of ranking scores from multiple sources can be very useful: combines multiple lists created using (independent) content representation, e.g. alternative ASR systems/methods.
- “Learning-to-rank” methods can be applied - importance of signals or alternative signals may be relevant - I’m not aware of SCR work that has explored this in practice.
- Other ranking features developed for text IR can be applied for SCR: non-content features such as: PageRank. Google quality metrics such as “Panda”, freshness, query ambiguity, etc.

Overview

sfi Fondúireacht Eolaíochta Éireann Science Foundation Ireland
DCU
CNGL Centre for Next Generation Localisation

Challenges for SCR

To develop effective SCR systems, it is important to go beyond naive combination of ASR and IR, and to address the following challanges.

Challenges for SCR

To develop effective SCR systems, it is important to go beyond naive combination of ASR and IR, and to address the following challanges.

The challenge of handling uncertainty

- ASR systems generate confidence scores of the certainty of the recognition of individual words.

Challenges for SCR

To develop effective SCR systems, it is important to go beyond naive combination of ASR and IR, and to address the following challanges.

The challenge of handling uncertainty

- ASR systems generate confidence scores of the certainty of the recognition of individual words.
- Challenge is how to determine appropriate ways to determine how to use this information in SCR.

Challenges for SCR

To develop effective SCR systems, it is important to go beyond naive combination of ASR and IR, and to address the following challanges.

The challenge of handling uncertainty

- ASR systems generate confidence scores of the certainty of the recognition of individual words.
- Challenge is how to determine appropriate ways to determine how to use this information in SCR.
- Challenge of how to use alternative unit (word) hypotheses to improve SCR.

Challenges for SCR

The challenge of covering all possible words

- ASR transcription systems are limited to a predefined fixed vocabulary.
- Words outside this vocabulary are referred to as “out-of-vocabulary” (OOV).

Challenges for SCR

The challenge of covering all possible words

- ASR transcription systems are limited to a predefined fixed vocabulary.
- Words outside this vocabulary are referred to as “out-of-vocabulary” (OOV).
- Using a larger vocabulary is not the answer:
  - Language is in constant growth and new words enter the vocabulary steadily.
  - Under certain circumstances, increasing the vocabulary can compromise ASR recognition accuracy.

Challenges for SCR

The challenge of covering all possible words

- ASR transcription systems are limited to a predefined fixed vocabulary.
- Words outside this vocabulary are referred to as “out-of-vocabulary” (OOV).
- Using a larger vocabulary is not the answer:
  - Language is in constant growth and new words enter the vocabulary steadily.
  - Under certain circumstances, increasing the vocabulary can compromise ASR recognition accuracy.
- Vocabulary can be adapted and new ASR system constructed, but
  - Can be difficult to obtain sufficient training data.
  - It is often not feasible to re-recognise spoken content - computational cost, collection size.

Challenges for SCR

The challenge of context

- Speech media is not fully represented by its transcript.
- User may require: certain speaker, speaking style, length of result, speech media of certain quality (no background noise).
- How should results be depicted in the user interface? How should speech segments be placed in context - be understandable in context, without requiring extensive audition.

Challenges for SCR

The challenge of structuring spoken content

- How should speech media be structured to support effective SCR?
- Much speech media does not have well-defined inherent structure.
- SCR systems must often carry out topical segmentation:
  - Select appropriate units for ranked IR.
  - How should content be represented to the user in the interactive interface?

Challenges for SCR

The challenge of visualization and playback

- Users are seldom willing to listen to long stretches of audio in order to access the information that they are looking for.
- A key feature of effective SCR is representing the content so that its relevance can be determine rapidly and accurately, and individual items can be interrogated efficiently.

Challenges for SCR

The challenge of visualization and playback

- Users are seldom willing to listen to long stretches of audio in order to access the information that they are looking for.
- A key feature of effective SCR is representing the content so that its relevance can be determine rapidly and accurately, and individual items can be interrogated efficiently.
- May include features of: graphical representation, recognized words, system confidence in words, topical or speaker involvement structure, suggested playback jump-in point(s).