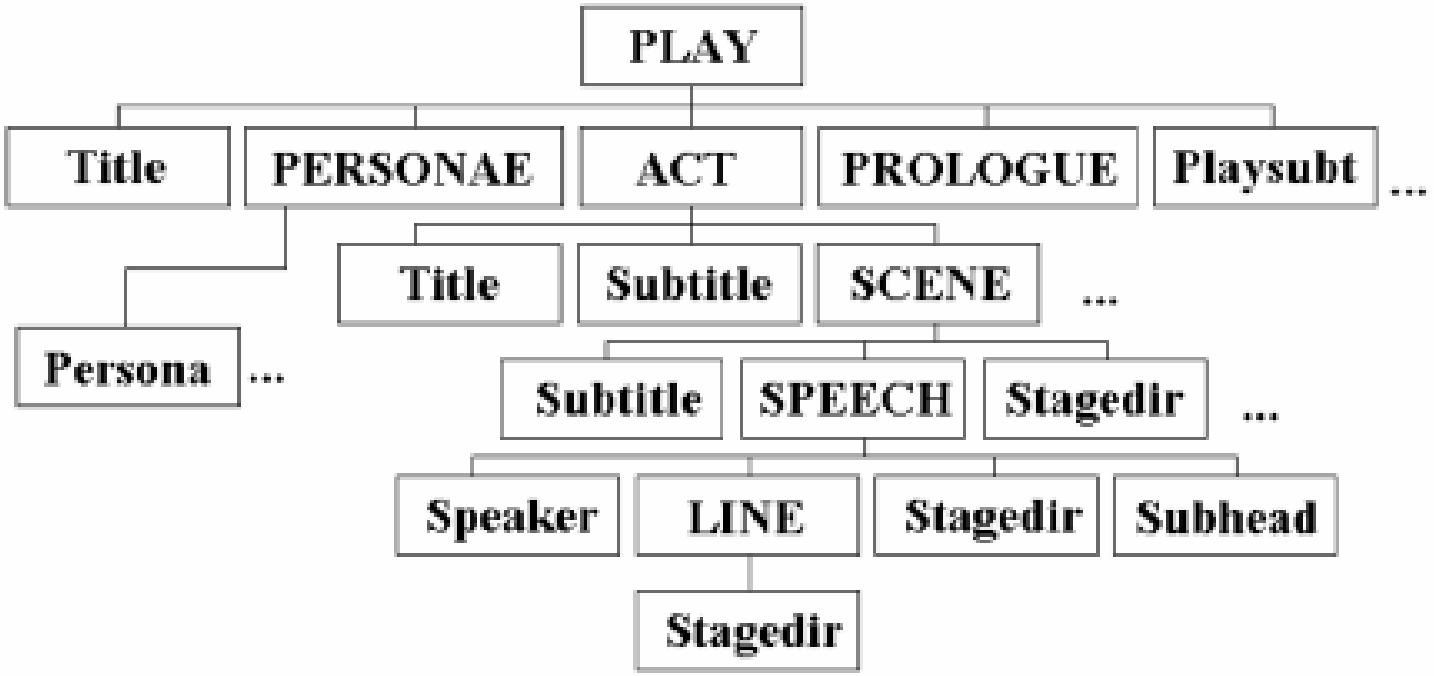
What does Shakespeare have to do with INEX?

Gabriella Kazai & Elham Ashoori

The Shakespeare user study

- Setup
  - 11 English and Drama students
  - 37 Shakespeare plays in XML
  - Subjects chose 3 plays each -> 12 plays selected
- Tasks
  - Create queries
  - Provide relevance assessments
  - Provide BEP assessments
- Focus of talk: Analysis of BEP selection strategies

Collection structure



Queries

- Formulate real information needs
  - No predefined topic/query format
  - No mention of possible query types: CAS vs. CO
  - Asked for varying complexity
    - Factual: "How old is Juliet?"
    - Essay topic: "The character of Lady Macbeth"
- 215 submitted queries
  - 43% CAS and 57 CO
  - Most common structural condition is PLAY! (80% of CAS) -> PLAY is treated as default contextual unit (support for Fetch & Browse task at INEX)

Relevance assessments

- Binary relevance and yellow-marker design
- 43 queries
- 2.7 assessor per query
- Highlighted text fragments then converted into leaf node level XML element judgements
- Assessor agreement
  - Optimistic relevance propagation to higher levels
  - 31% leaf, 35% SPEECH, 64% SCENE, 78% ACT, 100% PLAY -> judges agree on general area, but not exact location of relevant information

Assessment trends

- Assessment trend
  - Highlight only salient relevant sentences
  - Highlight whole context
- Depends on UI assumption
  - User will see relevant information as highlighted text within its context
  - User will be shown only the relevant information
- Q: INEX assessments UI vs. retrieval tasks?
  - Assessment UI matches Fetch & Browse task only

BEP assessments

- BEP = optimal starting point
- Obtained via interviews using a browsing UI and the union of the relevance assessments
- 521 unique BEPs (out of 928), 12 per query
- Assessor agreement
  - 49% leaf, 58% SPEECH, 51% SCENE, 0% PLAY

BEP selection strategies

- 94% BEPs at leaf or SPEECH level
  -> preference to more specific entry points
- BEP types:
  - Container BEP: parent node of relevant elements
  - "Start reading here" BEP: one node in a sequence of leaf nodes
  - Combined BEP: one node in a sequence of parent nodes

BEP selection strategies

- Start reading here BEP: 45%
  - 1st leaf node in a sequence: 62%
  - Other nodes: 10% (sometimes very last node) -> importance of seeing relevant answer then browse for context
  - 1 node sequence: 28%
- Container BEP: 31%
  - 80% SPEECH, 20% SCENE, ~0% PLAY
- Combined BEP: 24%
  - 1st SPEECH: 94%, middle: 6%

BEPs vs cluster of text

- Measure of specificity for BEP
  - Calculated as the number of relevant leaf nodes in cluster that is represented by a BEP/ total leaf nodes in cluster
  - E.g. Given cluster as 5r, 2i, 1r, 8i, 4r -> BEP spec = (5+1+4)/(5+2+1+8+4)
- Start reading here BEPs
  - 90% of the content is relevant -> most focused
- Container & combined BEPs
  - Only ~60% of the content is relevant

Conclusions

- Queries
  - CO and CAS naturally needed
  - Whole document as contextual natural semantic unit
- Relevance assessments
  - Agreement about the general area of relevance, but not necessary about the exact lines
- BEP assessments
  - Higher agreement than relevance assessments
  - Most specific is preferred (key relevant fragments)
  - Most popular BEP type is "Start reading here BEP" (these are also most specific)

Questions for discussion

- How do assumptions about the UI affect relevance assessments?
- How to convert highlighted passages to relevant XML elements?
- What is a BEP at INEX? Why just 1 BEP per doc?
- How to evaluate Best in Context? Distance measure?