

Transcript

[0:57][0:57] denis : here it is
[0:57][0:57] mirek : thank you
[0:58][0:59] andrei : i haven't
[0:59][1:27] denis : so the goal for today are
[0:59][1:27] denis : we have two goals uh
[0:59][1:27] denis : first is to decide a movie for uh 1

Preprocessing

- **Remove inutile characters:** ^, _, !, ...
- **Convert abbreviated words to official forms:**
I've -> I have,...
- **Convert numeric to text:** 34 -> thirty four, 2nd -> second, ...

Remove stop-words

a, about, an, are, as, at, am, and, be, by,
for, how, in, is, it, of, on, or, that, the,
they, this, to, so, uh, um, really, very, was,
were, we, well, will, with, wow

Word Splitting and Indexing

112 going go deni departure going leavi
proceed depart become get run pass lead
perish expire conk choke croak belong s
114 select select deni choose take choi
116 movie movi deni movie film picture

Position	Stemmer	Speaker	Synonyms
----------	---------	---------	----------

Pairs of two questions

96

Mirek nominated The Big Lebowski for the shc
Mirek nominated Usual Suspects for the short

97

Andrei nominates Saving Private Ryan
Andrei nominates Usual Suspects

Preprocessing

- **Remove inutile characters:** ^, _, !, ...
- **Convert abbreviated words to official forms:**
I've -> I have,...
- **Convert numeric to text:** 34 -> thirty four, 2nd -> second, ...

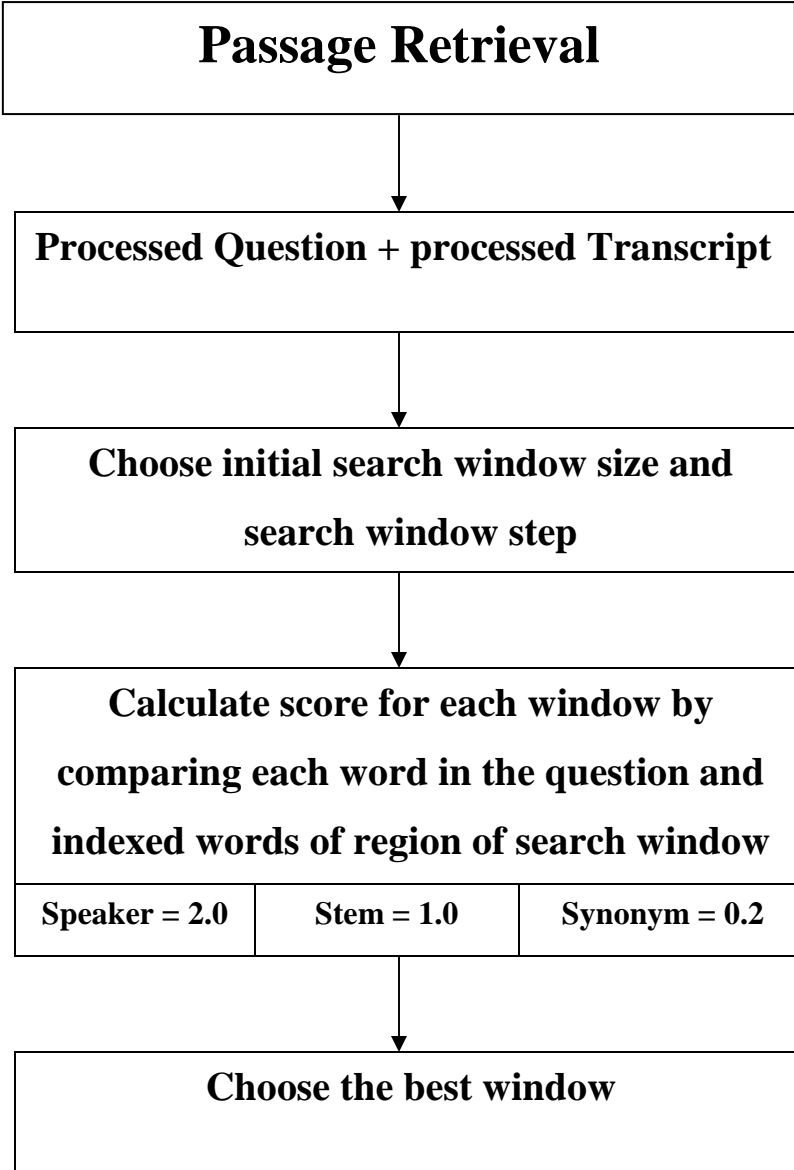
Remove stop-words

a, about, an, are, as, at, am, and, be, by,
for, how, in, is, it, of, on, or, that, the,
they, this, to, so, uh, um, really, very, was,
were, we, well, will, with, wow

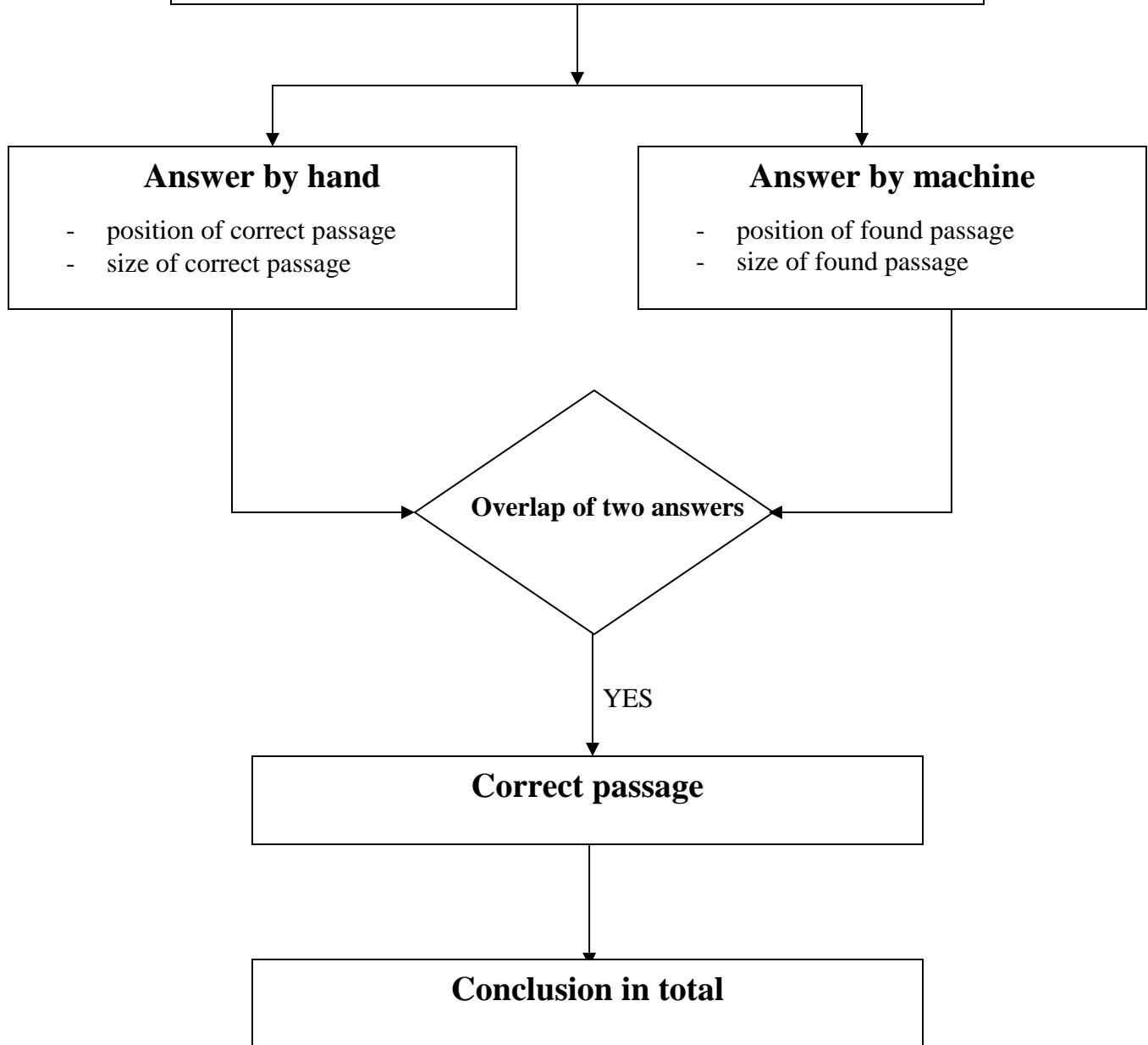
Word Lemmatization et Stem

Mirek had not received agenda meeting

- ⇒ mirek had/have not received/receive agenda meeting/meeting ././.
- ⇒ mirek had/have not receiv agenda meet ././.



Passage Retrieval Algorithm Evaluation



True-false question Answering

Pair of two statements

The first statement in the pair

The second statement in the pair

Passage Retrieval Algorithm

Passage Retrieval Algorithm

Retrieved Passage

Retrieved Passage

Score 1 ? Score 2

1st statement is true

2nd statement is true

Passage Analyze

