qn4_3

tokenization

- I
  -
- think
  -
- the
  -
- tricky
  -
- part
  -
- here
  -
- is
  -
- to
  -
- decide
  -
- when
  -
- single
  -
- quotes
  -
- are
  -
- parts
  -
- of
  -
- words
  -
- ,
  -
- when
  -
- periods
  -
- do
  -
- an

  -
- don't
  -
- imply
  -
- sentence
  -
- boundaries
  -
- ,
  -
- etc.
  -
- Sentence
  -
- splitting
  -
- is
  -
- a
  -
- deterministic
  -
- consequence
  -
- of
  -
- tokenization
  -
- :
  -
- a
  -
- sentence
  -
- ends
  -
- when
  -
- a
  -
- sentence-ending
  -
- character
  -
- (

  -
- .
  -
- ,
  -
- !
  -
- ,
  -
- or
  -
- ?
  -
-)
  -
- is
  -
- found
  -
- which
  -
- is
  -
- not
  -
- grouped
  -
- with
  -
- other
  -
- characters
  -
- into
  -
- a
  -
- token
  -
- (
  -
- such
  -
- as
  -
- for

  -
- an
  -
- abbreviation
  -
- or
  -
- number
  -
-)
  -
- ,
  -
- though
  -
- it
  -
- may
  -
- still
  -
- include
  -
- a few
  -
- tokens
  -
- that
  -
- can
  -
- follow
  -
- a
  -
- sentence
  -
- ending
  -
- character
  -
- as
  -
- part
  -
- of

  -
- the
  -
- same
  -
- sentence
  -
- (
  -
- such
  -
- as
  -
- quotes
  -
- and
  -
- brackets
  -
-)
  -
- .