Where are the Wikipedia pages associated with the questions?

For “how likely the answer is correct”, is there any sort of system that should be followed, ex. probability? Tf-idf Term-Weighting Approach?

* Problem 1 – Function that returns likeliness of answer being correct (Take in Wikipedia page and answer as input (ex. George W. Bush Page and Bush 43 and return how likely a certain answer is to be correct)
* Tf-idf on **all** words in the dataset of Wikipedia pages to determine weights
* Multiply weights by term frequency on given answer and add up
  + Answers Dataset
    - List of various dictionaries containing simulations of answers containing the following fields
      * Action
      * Date
      * Object
        + Room
        + User

ID

Name

* + - * + Playback Rate
        + Question Info

Category

Difficulty

Tournament

Num

Year

Round

* + - * + Question Text
        + Guess
        + Answer
        + Ruling
        + QID
        + Answer Duration
        + Time Elapsed
        + Time Remaining
  + Questions Dataset
    - Array of questions where each question contains the following fields
      * Text
      * Answer
      * Page
      * Category
      * Subcategory
      * Tournament
      * Difficulty
      * Year
      * Proto ID
      * QDB ID
      * Dataset
      * Qanta ID
      * Tokenization
      * First Sentence
      * Gameplay
      * Fold
* Problem 2 – Finding errors in Questions Dataset from above (Have a dictionary and detect which words are irregular or erroneous with the certain files)