SLP Exercise 2 August 25, 2020

**Link Analysis**

Read the first part of the Link Analysis chapter, pages 461-470, of Manning, Raghavan and Schutze, at <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf> .

1. Do any two of the three exercises at the bottom of page 463.
   * **Exercise 21.1**Is it always possible to follow directed edges (hyperlinks) in the web graph from any

node (web page) to any other? Why or why not?

No, because there are websites that actually do not point to any other website, or there are websites that used to connect to old, non-existent websites so the ‘flow’ is broken. Furthermore, for this to be true would mean that there is always a way to continue clicking hyperlinks, so in essence every website can be reached by any other website, but for example Dr.Kreinovich’s calendar does not contain hyperlinks (see: http://www.cs.utep.edu/vladik/09\_07\_20.html).

* + **Exercise 21.2**

Find an instance of misleading anchor-text on the Web.

This is found on: <https://theuselessweb.com>. But it is confusing, what if this implies, I can buy coffee from the content creator instead of buying a coffee for the creator? Furthermore, this is to give money to the creator which serves a different purpose than what you could think at first since the content creator might not actually receive/buy a coffee ever.

A screenshot of a cell phone

Description automatically generated

* + **Exercise 21.3**

Given the collection of anchor-text phrases for a web page *x*, suggest a heuristic for choosing one term or phrase from this collection that is most descriptive of *x*.

Heuristic:

1. For every term associated with webpage “X” compute the repetition of every term and add the term and it repetitions number to a list “A” as ordered pairs (term, repetition#).
2. For every pair of list “A” compute the score uniqueness(term)/repetition# and store into list “B” as (term, repetition#, uniqueness(term)/repetition#) ordered pairs.
   * + Where uniqueness(term) is a pre-computed score based on use of words throughout the internet.
3. Select max score in the list that contains pairs as: (term, repetition#, uniqueness(term)/repetition#).

\*Alpha = probability of teleportation

1. Write the transition probability matrix for Fig 21.2 on page 464, assuming alpha is 0.0.
   1. (Simple Hyperlink Mapping)
   2. (No teleportation and 1/N probability by row)
   3. (With teleportation)
2. Write a new transition probability matrix for the same figure, now assuming alpha is 0.1.
   1. Same steps as problem b

4) =

5) =

1. Roughly estimate the steady-state probability for this graph, again with alpha 0.1. Show your work.
   * P=

   * Approximate Values (of vector x after iteration 5):
     + P:[0.2036,0.2653,0.2654,0.2654]
   * Checked with a computer program:
     + P:[ 0.20408163265306137 0.2653061224489798 0.2653061224489798 0.2653061224489798]
   * Notes:

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated