Assignment 3

12.12.23

Maria Bruno (3636989), Francesca Carlon (3664579), Linnet Moxon (3387157)

Task 1

Subtask 1.1

	d1	d2	d3
street	0.23	0	0.18
bicycles	0.48	0	0
ride	0	0	0
on	0	0	0
horses	0	0.48	0
beach	0	0.48	0
bikes	0	0	0.48

computations:

$w_{t,d} = (1+\log_{10}(tf_{t,d})*\log_{10}(N/df_t)$

$w_{street,d1} = (1+\log_{10}(2))*\log_{10}(3/2) = 0.23$

$w_{street,d2} = 0$

$w_{street,d3} = (1+\log_{10}(1))*\log_{10}(3/2) = 0.18$

$w_{bicycles,d1} = (1+\log_{10}(1))*\log_{10}(3/1) = 0.48$

$w_{bicycles,d2} = 0$

$w_{bicycles,d3} = 0$

$w_{ride,d1} = (1+\log_{10}(1))*\log_{10}(3/3) = 0$

$w_{ride,d2} = (1+\log_{10}(1))*\log_{10}(3/3) = 0$

$w_{ride,d3} = (1+\log_{10}(1))*\log_{10}(3/3) = 0$

$w_{on,d1} = (1+\log_{10}(1))*\log_{10}(3/3) = 0$

$w_{on,d2} = (1+\log_{10}(1))*\log_{10}(3/3) = 0$

$w_{on,d3} = (1+\log_{10}(1))*\log_{10}(3/3) = 0$

$w_{horses,d1} = 0$

$w_{horses,d2} = (1+\log_{10}(1))*\log_{10}(3/1) = 0.48$

$w_{horses,d3} = 0$

$w_{beach,\ d1} = 0$

$w_{beach,d2} = (1+\log_{10}(1))*\log_{10}(3/1) = 0.48$

$w_{beach,d3} = 0$

$w_{bikes,d1} = 0$

$w_{bikes,\ d2} = 0$

$w_{bikes,d3} = (1+\log_{10}(1))*\log_{10}(3/1) = 0.48$

vectors for the documents:

v(d1) = [0.23 0.48 0 0 0 0 0]

v(d2) = [0 0 0 0 0.48 0.48 0]

v(d3) = [0.18 0 0 0 0 0 0.48]

Subtask 1.2

q: “bikes beaches” → we decided to apply stemming. Therefore, for this specific query, we will consider all the terms starting with “ bike* ” and “ beach* ”.

query vector: v(q) = [0.48, 0.48]

cos(q, d1) = ((0.48*0) + (0.48*0))/(sqrt(0.23^2 + 0.48^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2)*sqrt(0.48^2+0.48^2)) = 0

cos(q, d2) = ((0.48*0) + (0.48*0.48))/(sqrt(0^2 + 0^2 + 0^2 + 0^2 + 0.48^2 + 0.48^2 + 0^2)*sqrt(0.48^2+0.48^2)) = 0.42

cos(q, d3) = ((0.48*0.48) + (0.48*0))/(sqrt(0.18^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0.48^2)*sqrt(0.48^2+0.48^2)) = 0.19

ranking:

1. doc2
2. doc3
3. doc1

Task 2

Step 0:

“ I want to find vegan dessert recipes that use cherries and almonds, but I'm allergic to hazelnuts”.

query → vegan dessert with cherries almonds no hazelnuts

Step 1:

vegan dessert with cherries almonds no hazelnuts

Show more

One Green Planet
https://www.onegreenplanet.org › vegan-recipe › nuss...

German Hazelnut and Almond Cake With Cherries [Vegan]

Ingredients You Need for Nusskuchen: German **Hazelnut** and **Almond Cake With Cherries** [**Vegan**] · 1 1/2 cups semolina · 1 1/4 cups ground **hazelnuts** · 4.4 ounces ground ...

The Vegan Chef School
https://theveganchefschool.com › recipeblog › cherry-...

Vegan Cherry Bomb Recipe

1 Jul 2021 — ... **CHERRIES**. 6 fresh **cherries**. 6 **hazelnuts** t and g. 100g sugar. FLUFF. 75ml (5T) aquafaba. Few drops **almond** essence. 1/4t vanilla paste. 1/2t cream ...

Pacific Foods
https://www.pacificfoods.com › Recipes

Vegan Cherry Almond Ice Cream Pops

Vegan Almond Butter Cookie Skillet · Chocolate Mini Cakes with **Hazelnut** Cream · **Vegan** Vanilla Cranberry Cupcakes · **Vegan Almond** Butter Cookie Skillet · Chocolate ...
6 hrs 10 mins

vegan dessert with cherries almonds no hazelnuts

Bianca Zapatka
https://biancazapatka.com › nut-cream-cake

Nut Cream Cake - Vegan Layer Cake Recipe

17 Dec 2022 — Step 1: Make the **hazelnut cake** layers. Start by sifting flour and baking powder into a large bowl. Add ground **hazelnuts**, baking soda, salt, ...
★★★★★ Rating: 4,8 · 5 votes · 1 hr 20 mins

One Green Planet
https://www.onegreenplanet.org › vegan-food › vegan...

15 Vegan Cherry Recipes to Make Before the Season is ...

This Nusskuchen: German **Hazelnut** and **Almond Cake With Cherries** by Peffe Stahl is an excellent **dessert** for any holiday occasion. 12. Raw **Cherry** Cupcakes With ...

dailysupertasty.com
https://www.dailysupertasty.com › vegan-cherry-hazel...

Vegan Cherry Hazelnut Cake

19 Mar 2023 — The perfect **cake** for the weekend! **Vegan cherry hazelnut cake**. With juicy base and delicious **cherry** cream topping.

The Plant Based School
https://theplantbasedschool.com › Desserts

Biscotti Recipe

23 Aug 2023 — Substitute **hazelnuts**, **walnuts**, pecan, dark chocolate chips, pistachios, raisins, dried **cherries** or cranberries. Salt: A pinch of sea salt ...
★★★★★ Rating: 4,9 · 61 votes · 1 hr

Bianca Zapatka
https://biancazapatka.com › vegan-baked-oats

Vegan Baked Oats Recipe | Baked Oatmeal

22 Jan 2022 — ... **cherries** are great too! Same goes for the **almonds** – Feel free to use other chopped **nuts** like **hazelnuts**, **walnuts**, cashews or whatever you like!
★★★★★ Rating: 4,8 · 8 votes · 25 mins

vegan dessert with cherries almonds no hazelnuts

Minimalist Baker
https://minimalistbaker.com › tag › cherry

cherry Archives

Maple **Almond** Buckwheat Granola · **Cherries** and **hazelnuts** beside a bowl of our roasted beet and **cherry** salad. Roasted Beet & **Cherry** Salad with Balsamic ...

Yummly
https://www.yummly.com › recipes › almond-milk-desserts

10 Best Almond Milk Desserts Recipes

mixed fruit, **cherries**, **vegan** ... **Hazelnut Cake Recipe**. One Bowl Cinnamon, Apple + **Hazelnut CakeGoodness** Green. 7. **almond** milk, chopped **hazelnuts**, apple, chia seed ...

Lauren Fit Foodie
https://laurenfitfoodie.com › ... › Frostings & Dips

Healthy Nutella Recipe (Half the Calories + No Hazelnuts ...

20 Jun 2022 — Ingredients and Substitutions · Nut butter – **Hazelnut** butter would work best since Nutella is made from **hazelnuts**. · Cocoa powder · Pancake syrup – ...

★★★★★ Rating: 5 · 5 votes · 5 mins

bloomingwithflavors.com
https://bloomingwithflavors.com › cherry-crumble

Vegan Cherry Crumble with oats and poppy seeds

7 Aug 2021 — This **vegan cherry** crumble with ground **hazelnuts**, oat flakes, butter ... Raspberry **almond cake** with **no** egg (**vegan**). Coconut yogurt with ...

★★★★★ Rating: 5 · 2 reviews

A Couple Cooks
https://www.acouplecooks.com › cherry-hazelnut-ener...

Cherry Vegan Energy Bars

14 May 2017 — 2 cups **hazelnuts** · 20 Medjool dates (2 cups) · 1 cup + ¼ cup tart dried **cherries** (with **no** sugar added), divided · Pinch kosher salt · ½ teaspoon ...

★★★★★ Rating: 5 · 1 review · 1 hr

Resplendent Kitchen
https://www.resplendentkitchen.com › Recipes › Dessert

Vegan Linzer Cookies with Almond and Hazelnut

19 Dec 2019 — These **hazelnut** and **almond Vegan** Linzer Cookies are a classic cut-out cookie during the holiday season. These jam-filled cookies are filled with

vegan dessert with cherries almonds no hazelnuts

Addicted to Dates
https://addictedtodates.com › ... › Brownies and Slices

No Bake Date Brownies

19 Jan 2023 — You can use pecans if you don't have **walnuts** to hand. **Hazelnut** butter: If you're **not** a fan of **hazelnuts** you can make these with any nut butter ...

★★★★★ Rating: 5 · 17 votes · 2 hrs 20 mins

delicious.com.au
https://www.delicious.com.au › gallery › tqrzkdaf

26 gluten-free desserts to ring in the weekend - Delicious

Hazelnut chocolate dacquoise **cake** with red wine **cherries** · Spiced carrot **cake** with ricotta icing · Matt Preston's golden **almond**, pear and raspberry **cake** · Gluten- ...

Tasty Thrifty Timely
https://tastythriftytimely.com › Recipes › Dessert

Simple Vegan Chocolate Mousse Terrarium For Two

4 Feb 2022 — A **vegan** chocolate mousse **dessert** with decadent avocado mousse, **cherry** sauce, and toasted **hazelnut** cookie crumble ... ¼ cup **hazelnuts**, toasted (sub ...

30 mins

Ottolenghi
https://ottolenghi.co.uk › recipes › cakes-desserts

Cakes & Dessert Recipes

Danish rice pudding (Risalamande) with salted **almond** brittle and sour sumac **cherries** ... **hazelnuts** and orange oil · READ MORE. Funky faloodeh · READ MORE ...

RecipeTin Eats
https://www.recipetineats.com › Cakes

Hazelnut Olive Oil Cake (GF, DF)

16 Sept 2022 — 1. **Hazelnut** meal is ground **hazelnuts**. It's like **almond** meal except, well, made using **hazelnuts**! Find it in the dried fruit and nut section of ...

★★★★★ Rating: 4,8 · 31 votes · 1 hr

fabsgoodfood.com
https://fabsgoodfood.com › Recetas

Jam Cake with toasted Hazelnuts (vegan) - Fabs Good Food

What we did: we used Google as our search engine. These are the screenshots for the top 20 results for the query “vegan dessert with cherries almonds no hazelnuts”.

Step 2:

		Maria	Francesca	Linnet
1	https://www.onegreenplanet.org/vegan-recipe/nusskuchen-german-hazelnut-and-almond-cake-with-cherries/	Not relevant	Not relevant	not relevant
2	https://theveganchefschool.com/reciepeblog/cherry-bomb	Not relevant	Not relevant	not relevant
3	https://www.pacificfoods.com/recipes/vegan-cherry-almond-ice-cream-pops/	Not relevant	Relevant	relevant
4	https://biancazapatka.com/en/nut-cream-cake/	Not relevant	Not relevant	not relevant
5	https://www.onegreenplanet.org/vegan-recipe/black-bottom-cherry-pie/	Relevant	Relevant	relevant
6	https://www.dailysupertasty.com/vegan-cherry-hazelnut-cake/	Not relevant	Not relevant	not relevant
7	https://theplantbasedschool.com/almond-biscotti/	Relevant	Relevant	not relevant
8	https://biancazapatka.com/en/vegan-baked-oats/	Relevant	Relevant	relevant
9	https://minimalistbaker.com/maple-almond-buckwheat-granola/	Relevant	Not relevant	relevant
10	https://www.yummly.com/recipes/almond-milk-desserts	Not relevant	Not relevant	not relevant
11	https://laurenfitfoodie.com/homemade-nutella-recipe/	Not relevant	Not relevant	not relevant
12	https://bloomingwithflavors.com/cherry-crumble/	Not relevant	Not relevant	relevant
13	https://www.acouplecooks.com/cherry-hazelnut-energy-bars/	Relevant	Relevant	not relevant
14	https://www.resplendentkitchen.com/vegan-linzer-cookies/	Not relevant	Not relevant	not relevant
15	https://addictedtodates.com/vegan-nutella-brownies-no-bake/	Not relevant	Not relevant	not relevant
16	https://www.delicious.com.au/recipes/collections/gallery/26-gluten-free-desserts-to-ring-in-the-weekend/tqrzkdaf?page=23	not relevant	relevant	not relevant

17	https://tastythriftytimely.com/simple-vegan-chocolate-mousse-terrarium-for-two/	relevant	relevant	relevant
18	https://ottolenghi.co.uk/recipes/cakes-desserts	relevant	relevant	not relevant
19	https://www.recipetineats.com/charlottes-hazelnut-olive-oil-cake/	Not relevant	Not relevant	not relevant
20	https://fabsgoodfood.com/recipe/jam-cake-with-toasted-hazelnuts-vegan/	Not relevant	Not relevant	relevant

What we did: we copied and pasted the links to the web pages from the top 20 results in a table. Then, separately, we opened each of them and we annotated with the labels “relevant” and “not relevant” according to whether the result was consistent with our query.

Step 3:

Cohen’s kappa Maria - Francesca

	Relevant(Maria)	Not relevant(Maria)
Relevant(Francesca)	6	2
Not relevant(Francesca)	1	11

P(A) = (6 + 11) / 20 = 0.85

P(E) = 7/20 * 8/20 + 13/20 * 12/20 = 0.14 + 0.39 = 0.53

Cohen's k = (0.85 - 0.53) / 1 - 0.53 = 0.32 / 0.47 = 0.68

Cohen’s kappa Maria - Linnet

	Relevant(Maria)	Not relevant(Maria)
Relevant(Linnet)	4	3
Not relevant(Linnet)	3	10

P(A) = (4+10)/20 = 0.7

P(E) = 7/20 * 7/20 + 13/20 * 13/20 = 0.545

Cohen’s k = (0.7-0.545)/(1-0.545) = 0.155/0.455 = 0.34

Cohen’s kappa Linnet - Francesca

	Relevant(Linnet)	Not relevant(Linnet)
Relevant(Francesca)	4	4
Not relevant(Francesca)	3	9

P(A) = (4+9)/20 = 0.65

P(E) = 7/20 * 8/20 + 12/20 * 13/20 = 0.53

Cohen’s k = (0.65-0.53)/(1-0.53) = 0.12/0.47 = 0.26

Step 4:

We annotated differently in the cases where we had different interpretations about the relevant documents. For example, some of us did not agree when the link directed us to a page full of recipes to choose from, where the request for the ingredients (cherry, almond but not hazelnuts) would appear even in separate recipes. Another case could be when the recipe requires an ingredient in another “shape”: for example almond butter instead of almonds. Another aspect for disagreement was, when the recipe said that the hazelnuts or for example the blueberries can be easily substituted by other nuts or fruits.

We observed that there was a high agreement between two annotators, but a lower agreement between the other two pairs of annotators.

Maria and Francesca (k = 0.68) had substantial agreement.

Maria and Linnet (k = 0.34) and Francesca and Linnet (k = 0.26) had a fair agreement.

Regarding the qualitative analysis about the disagreement we had, as mentioned above, we disagreed in the cases in which the link in the results page led to a webpage containing links to different recipes. While Linnet and Francesca annotated that link as not relevant, Maria annotated it as relevant if she could find a relevant recipe among one of the ones proposed.

Also, in cases 7 and 13 Linnet marked both results as not relevant, although the recipe shows substituting variations with cherries or other nuts (which she would consider relevant in other cases), because she didn’t read the page to the very end - opposed to Francesca and Maria. There is a similarity to the page that lists other pages, where the willingness to read everything to the end or to go to other pages was not there.

Here you can see that for “personal” ranking not only the plain content of the page is relevant but also the structure of the page and where the relevant information can be found.

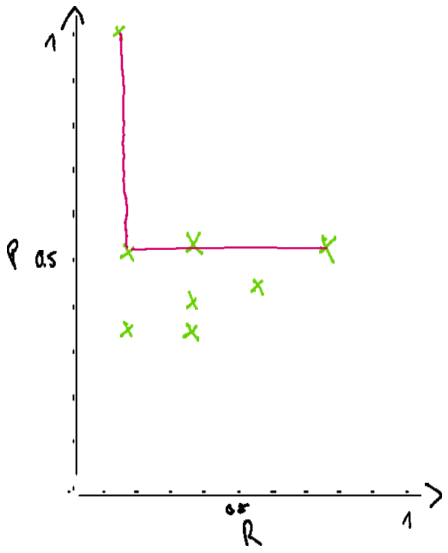
Task 3

precision = (retrieved&relevant)/((retrieved&relevant)+(retrieved&non-relevant))

recall = (retrieved&relevant)/((retrieved&relevant)+(not-retrieved&relevant))

k	result set	precision	recall
1	128	1	0.2
2	128, 7	0.5	0.2
3	128, 7, 9	1/3	0.2
4	128, 7, 9, 3	0.5	0.4
5	128, 7, 9, 3, 35	0.4	0.4
6	128, 7, 9, 3, 35, 32	1/3	0.4
7	128, 7, 9, 3, 35, 32, 41	3/7	0.6
8	128, 7, 9, 3, 35, 32, 41, 64	0.5	0.8

Interpolation graph:



Task 4

We discussed in the first lecture of this class how we can efficiently create an inverted index for document-at-a-time processing. We did not discuss in detail how to create an index for term-at-a-time processing.

Please develop an efficient strategy to do so and explain it.

In order to build a term-at-a-time index, we can create a list of the most relevant words appearing in the documents provided, together with the docID where that term appears in and its weight for each docID. So, potentially we would create something like:

Term	**docID, weight of the term**
HARRY	1, 0.3 ; 3, 0.04 ; 5, 0.66 …
HERMIONE	1, 0.4 ; 2, 0.002, ; 3, 0.05 …
RON	2, 0.06 ; 3, 0.05 ; 5, 0.63 …

Then, let's consider that our query is HARRY, HERMIONE.

first, we process the weight for HARRY:

this terms appears in doc1 with a weight of 0.3, and so on with the other documents and weights so let's write:

- 1: 0.3
- 3: 0.04
- 5: 0.66

now let's go over HERMIONE by building up from HARRY in case they appear in the same document:

- 1: 0.3 + 0.4 = 0.7
- 2: 0.002
- 3: 0.04 + 0.05 = 0.09

As a parameter, we might decide that if the weight of that term in the document is not high enough (i.e. not so relevant), we can decide not to consider it and therefore not include it in the query process. We might then set the threshold as follows: exclude weight <= 0.001 in the counting for ranking creation.

Now we can proceed to the ranking results: we take the highest weights and we rank them.

1. Doc1
2. Doc5

3. Doc3
4. Doc2

In conclusion, for the query HARRY, HERMIONE the most relevant document is Doc1, followed by doc5, doc3 and doc2.

Further, explain the differences between the two creation procedures. Which of the two indexing procedures is more efficiently done?

The **document-at-a-time** procedure allows the retrieval of relevant information by creating an inverted index with docID and the words within:

- Doc1 → Stuttgart, Germany, Europe…
- Doc2 → Bavaria, Munich, Germany…
- Doc3 → Germany, Baden-Württemberg, Bavaria…

The query for [Bavaria, Germany] will result in the set intersection of documents {2, 3}.

On the other hand, the **term-at-a-time** procedure sorts the relevant terms out according to docID and its weight within that document (we can decide for a threshold to consider the term relevant or not as indicated before). See example for HARRY, HERMIONE, RON in the previous answer.

The query will then make use of the weights to sum them up if there is a correspondence between their presence in the documents. Finally, the results of the query (terms) will give us a ranking of the most relevant documents where the query terms have the most weight.

To sum up, the main differences between the two procedures are:

- the presence of weights in term-at-a-time
- more potential parameters in term-at-a-time
- the difference in results
    - *set of documents* vs. *ranking of documents*

The most efficient procedure might differ according to the task, however the term-at-a-time one might make the user happier because it does not only give a final list of the relevant documents, but it also ranks them in order of relevance. Therefore, the user is likely to find relevant information more quickly because they might follow an order instead of randomly jumping from a document to another.

PROGRAMMING TASK

Subtask 2

```
import numpy as np


"""
function to compute tf
has to be done separately for each tweet
:return a dictionary with the term and its tf: {'term': tf, 'term': tf, ...}
:parameter one tweet
"""



def compute_tf(text):
    term = 0
    vector = []

# calculate frequency of each word in text

    words = []
    words = text.split()
    freq_dict = {}

# loop through words in text and make a dictionary for word, frequency

    for w in words:
        word_freq = words.count(w)
        # print(word_freq)
```

```
        freq_dict[w] = word_freq


# calculate term frequency for word in text
    total_words = len(words)
    frequency = 0
    # print(total_words)


    tf_dict = {}
    for word in words:
        for k in freq_dict.keys():
            if k == word:
                frequency = freq_dict[k]


        term_frequency = frequency / total_words
        tf_dict[word] = term_frequency
        # print(term_frequency)
    # print(tf_dict)


    return tf_dict



"""
function to compute tf (number of documents that a term occurs in)
done for both tweets/documents together
:return a dictionary with each term and its idf: {'term': idf, 'term': idf, ...}
:parameter both tweets
"""
```

```
def compute_idf(doc1, doc2):
    # total number of docs is 2
    tot_docs = 2
    idf_dict = {}

    # join the two input tweets into one string
    total_string = doc1 + " " + doc2
    total_string = total_string.split()

    # create a dictionary with the idf per term
    for t in total_string:
        tot_word_freq = total_string.count(t)
        idf_dict[t] = tot_docs/tot_word_freq

    return idf_dict



"""
function to compute the weights per term (w =
(1+log(tf))*log(idf))
:return two vectors with weights: one for each tweet
:parameter both tweets
"""



def compute_weights(doc1, doc2):
```

```
    vector_doc1 = []
    vector_doc2 = []
    tf1 = compute_tf(doc1)
    tf2 = compute_tf(doc2)
    idf = compute_idf(doc1, doc2)

    # go through doc1 and compute weight for every term

    for v in tf1.values():
        for n in idf.values():
            v = int(v)
            n = int(n)

            weight = (1+np.log10(v))*(np.log10(n))
            vector_doc1.append(weight)

    # go through doc2 and compute weight for every term

    return vector_doc1, vector_doc2



tweet_one = "says lifetime risk of cervical cancer in Japan is 1 in 100.  That means HPV is endemic in Japan, and screening is not working well."
tweet_two = "New Study Links Cellphone Radiation to Heart and Brain Tumors"
#print(compute_tf(tweet_one))
print(compute_weights(tweet_one, tweet_two))
```

```
'''The next steps to finish this part of the exercise would be, first, adding the correct

weights to the vectors for the documents. The same method that was applied to create

the vector for document 1 should be applied to create the vector for document 2.

Then, a new function should be created to calculate the cosine similarity between the vectors.'''
```