Recipe Recommendation

Yuzhou Tu
Research
August 30, 2018

PHILIPS

Project: Recipe Recommendation

PHILIPS Kitchen Applicance

Recipe Recommendation

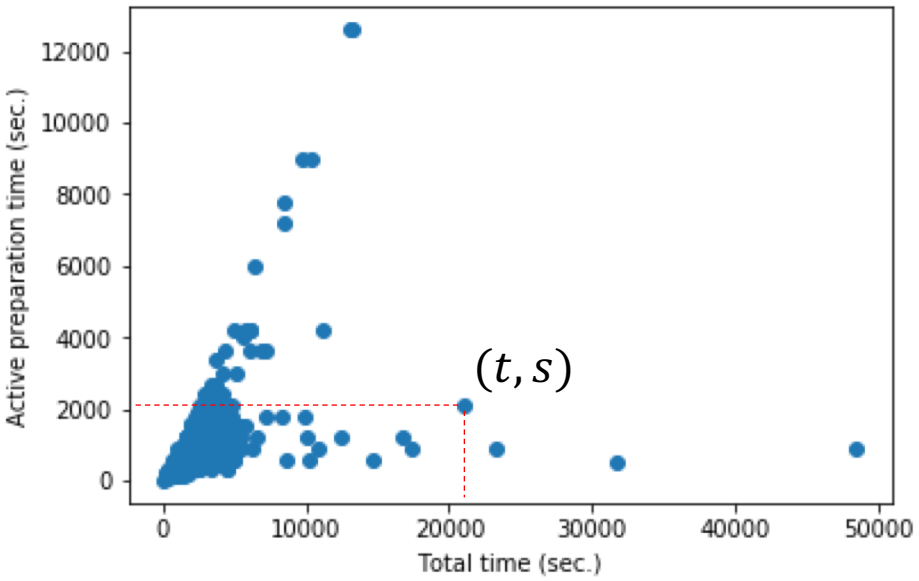
- ➢ Difficulty
  - Time
  - Steps
  - Ingredients
- ➢ Flavor/ Ingredients
  - Main Ingredients
  - Seasonings
  - Overall

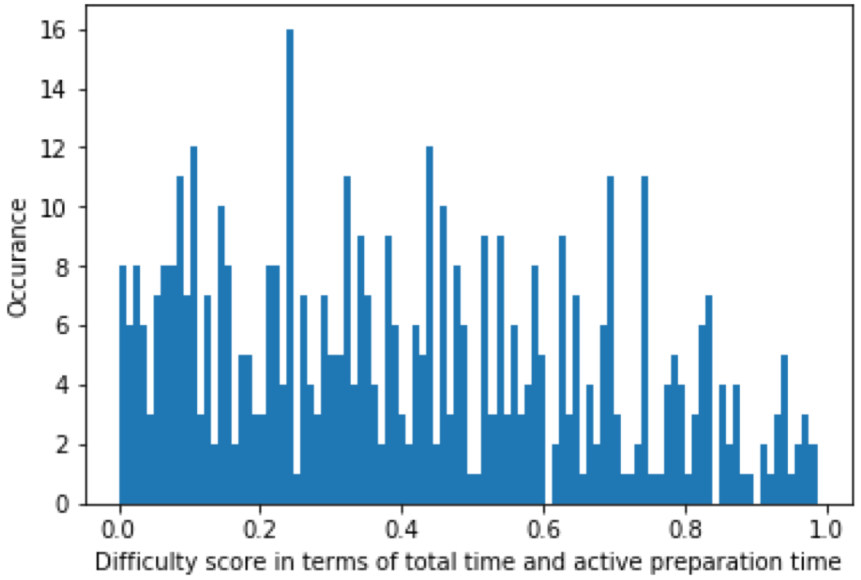
PHILIPS

Difficulty Score by Time: Total & Active Preparation Time

- Difficulty score for a recipe with total time t and active preparation time s:

$$\text{score}(t,s) = \frac{\#\{r \mid \text{recipe } r\text{'s total time} \leq t \text{ and active preparation time} \leq s\}}{\#\text{all recipes}}$$



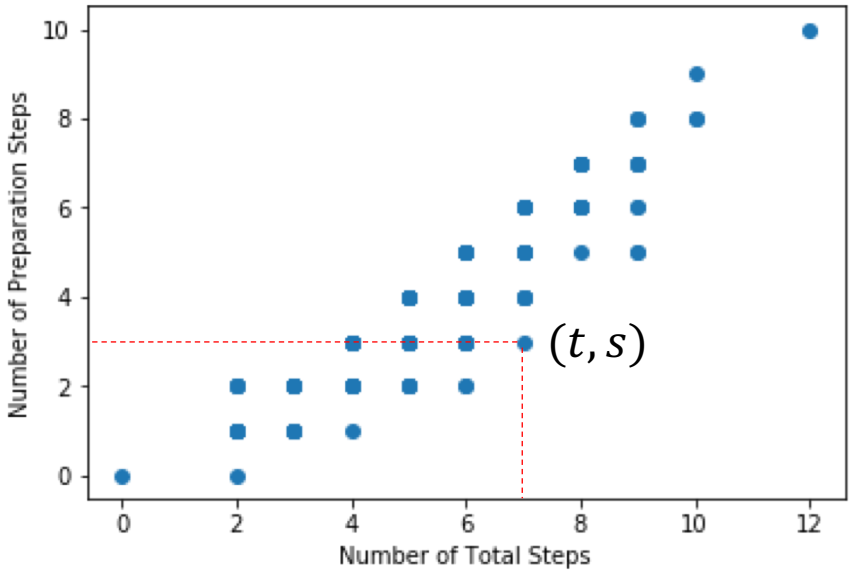


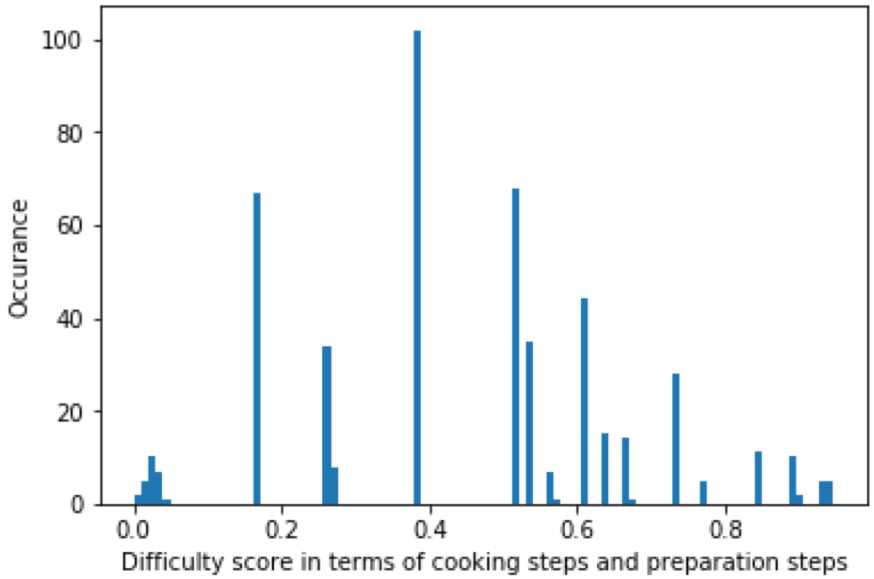
PHILIPS

Difficulty by Recipe Steps: Total & Preparation Steps

- Difficulty score for a recipe with total steps t and preparation steps s:

$$\text{score}(t,s) = \frac{\#\{r | \text{recipe } r\text{'s total steps} \leq t \text{ and preparation steps} \leq s\}}{\#\text{all recipes}}$$





PHILIPS

Difficulty Score by Ingredients: Number & Availability

- Availability of an ingredient i is related to its probability $P(i) = \frac{\#\{r | i \text{ in recipe } r\}}{\#\text{all recipes}}$:
  - Higher $P(i)$ indicates higher availability of such ingredient i, and vice versa
- The difficulty of an ingredient is thus evaluated by $-\log\big(P(i)\big) \geq 0$
- The difficulty index diff_idx of a recipe r is the summation of $-\log\big(P(i)\big)$ for all ingredient i in r, i.e.,

$$\text{diff\_idx}(r) = \sum_{i \text{ in } r} -\log\big(P(i)\big)$$

- Use probability to rescale the difficulty index into a score between 0 and 1:

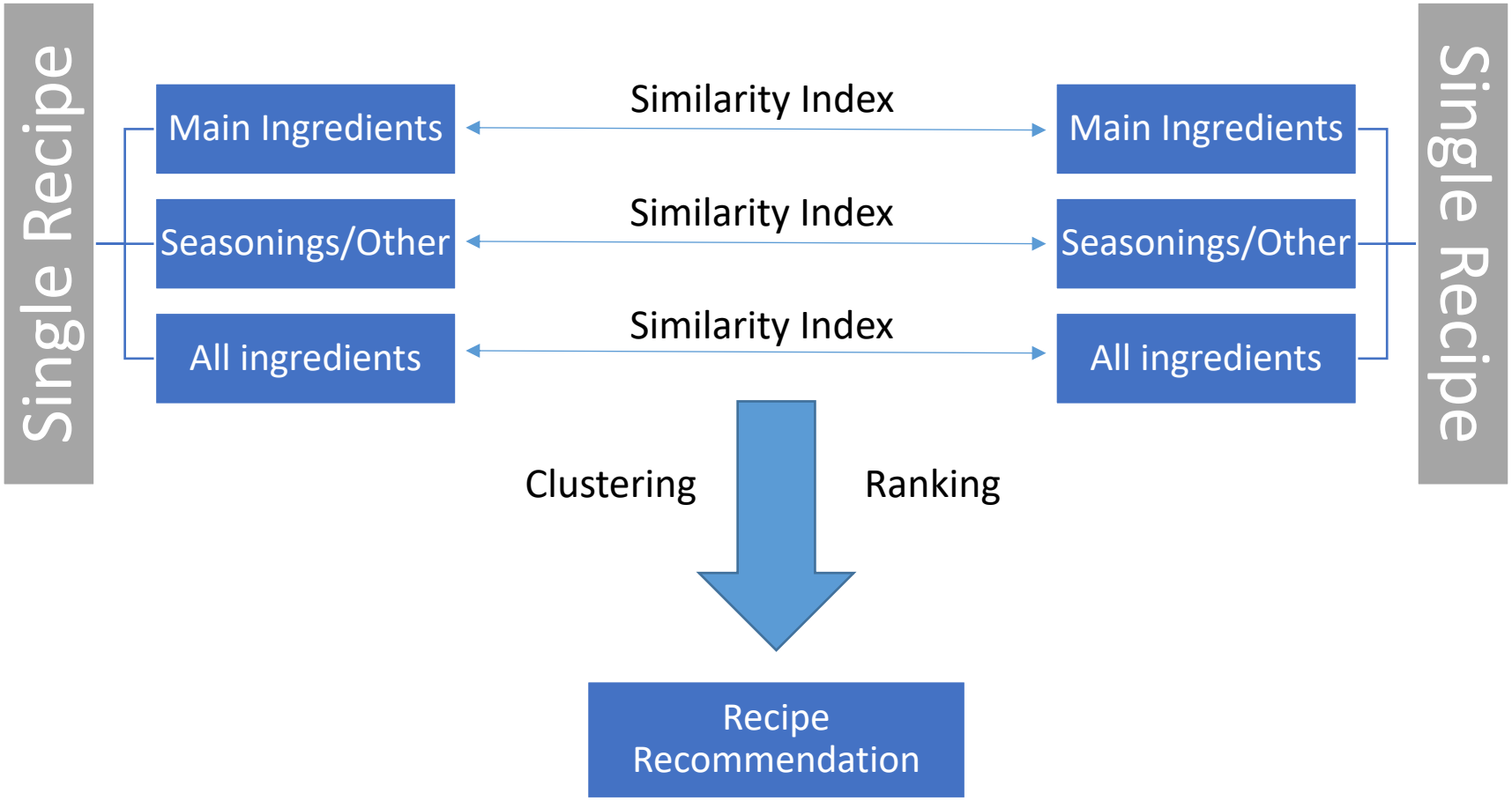
$$\text{score}(r) = \frac{\#\{r' | \text{diff\_idx}(r') \leq \text{diff\_idx}(r)\}}{\#\text{all recipes}}$$

PHILIPS

Recipe Recommendation – Flavor/ Ingredients

- Determination of Main Ingredients and Seasonings
- Similarity Index
- Clustering

Recipe Recommendation



PHILIPS

Main Ingredient Determination

Higher Weight Fraction

- Weight fraction of ingredient i in the recipe r: w_{ri}

Less Common

- Probability of Ingredients i: $P_i = \frac{\#\{r | i \text{ in recipe } r\}}{\#\text{all recipes}}$

$$Entropy = -\sum w_{ri} \cdot \log(p_i)$$

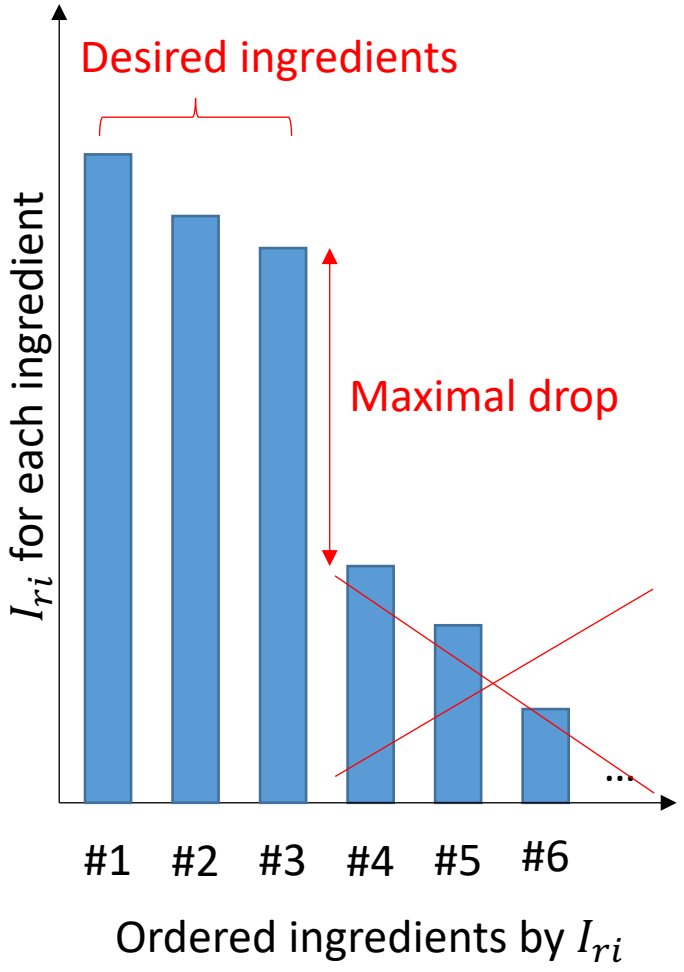
Main Ingredient Indicator

- Main ingredient indicator for ingredient i in the recipe r: $I_{ri} = w_{ri} \cdot (-\log(P_i))$

PHILIPS

Main Ingredient Determination

- Determine **main ingredients** by maximal drop of I_{ri}
  - Sort all ingredients in recipe r in descending order of I_{ri}, a.k.a., $I_{ri_1} \geq I_{ri_2} \geq \cdots \geq I_{ri_n} \geq I_{ri_{n+1}} \geq \cdots$
  - Find k so that $I_{ri_k} - I_{ri_{k+1}} = \max_n \{I_{ri_n} - I_{ri_{n+1}}\}$

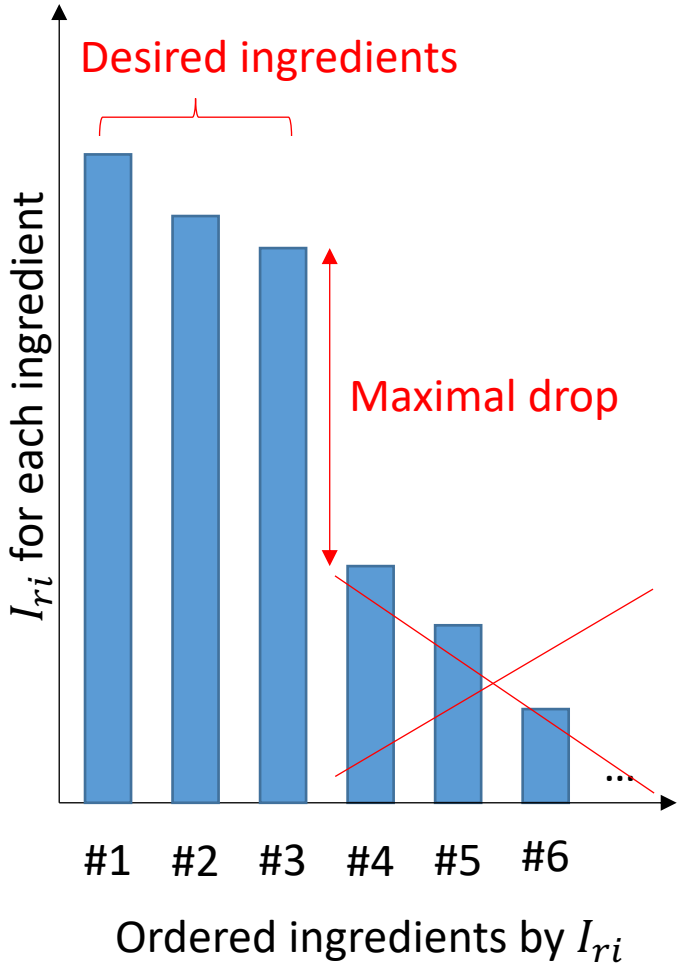


PHILIPS

Main Ingredient Determination

Crispy Potato Skin Wedges

	pepper	salt	Russet Potato	Paprika Powder	canola oil	water
weight (gram)	4.5	2.5	400	4.5	2.25	250
weight fraction (%)	0.6780	0.3766	60.2637	0.6780	0.3390	37.6648
log(1/p)	0.3696	0.1815	2.0065	1.1826	1.7054	0.9853
I	0.25061	0.06835	120.91697	0.80173	0.57811	37.11024



PHILIPS

Recipe Similarity Index based on Main Ingredients

- Find common main ingredients in recipe pairs
- Calculate Jaccard similarity

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

- Main ingredient similarity index between recipes r_1 and r_2:

$$\mathrm{Sim}(r_1, r_2) = \frac{\sum_{i \text{ is } common \text{ main ingr in } r_1 \text{ and } r_2} \min(I_{r_1 i}, I_{r_2 i})}{\sum_{i \text{ is main ingr in } r_1 \text{ or } r_2} \max(I_{r_1 i}, I_{r_2 i})}$$

$$= \frac{\sum_{i \text{ is } common\ main \text{ ingr in } r_1 \text{ and } r_2} \min(w_{r_1 i},\ w_{r_2 i}) \cdot \log(P_i)}{\sum_{i \text{ is } main\ ingr \text{ in } r_1 \text{ or } r_2} \max(w_{r_1 i},\ w_{r_2 i}) \cdot \log(P_i)}$$

PHILIPS

Recipe Similarity Index based on Flavor

- Hypothesis: Recipe flavor is strongly related to seasonings

  - Determine seasoning i in recipe r based on $I_{ri} = -(1 - w_{ri}) \cdot \log(1 - P_i)$

  - Seasoning similarity index between recipes r_1 and r_2:

$$\mathrm{Sim}(r_1, r_2) = \frac{\sum_{i \text{ is common seasonings in } r_1 \text{ and } r_2} \min(w_{r_1 i}\,,\ w_{r_2 i}) \cdot \log(1 - P_i)}{\sum_{i \text{ is seasonings in } r_1 \text{ or } r_2} \max(w_{r_1 i}\,,\ w_{r_2 i}) \cdot \log(1 - P_i)}$$

- Hypothesis: Recipe flavor is determined by all ingredients other than main ingredients

  - Determine flavor ingredients i in recipe r based on {All Ingredients} / {Main Ingredients}

  - Flavor similarity index between recipes r_1 and r_2:

$$\mathrm{Sim}(r_1, r_2) = \frac{\sum_{i \text{ is common flavor ingredients in } r_1 \text{ and } r_2} \min(w_{r_1 i}\,,\ w_{r_2 i}) \cdot \log(1 - P_i)}{\sum_{i \text{ is flavor ingredients in } r_1 \text{ or } r_2} \max(w_{r_1 i}\,,\ w_{r_2 i}) \cdot \log(1 - P_i)}$$

PHILIPS

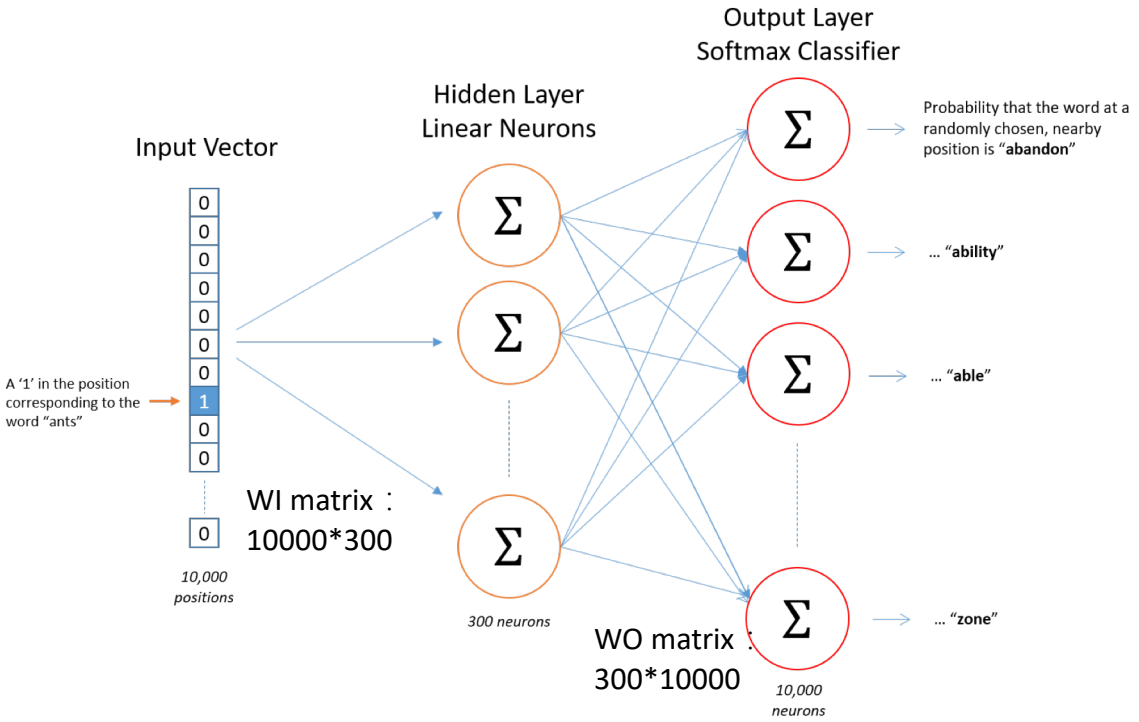
Further Improvement enabled by Word2Vec

- Problem:
  - In the previous similarity indices, the similarity between two ingredients i and j is not taken into account. For instance, similar ingredients beef, sirloin, and tenderloin are viewed as different entities in the previous calculation

- Word2Vec model enable us to digitalize word into vector, and it works for food ingredient as well
  - Each ingredient i can be vectorized as $\mathrm{vec}(i)$
  - If an ingredient contains several words, $\mathrm{vec}(i)$ will be the average vector of each constituent word
  - Similarity of the pair of ingredients, say, i and j, can be calculated as $\mathrm{Sim}(i,j) = \frac{\mathrm{vec}(i)\cdot\mathrm{vec}(j)}{\|\mathrm{vec}(i)\|\|\mathrm{vec}(j)\|}$
  - Google's trained Word2Vec model is applied for retrieving $\mathrm{vec}(i)$

PHILIPS

Word2Vec



WI =

-0.094491	-0.443977	0.313917
-0.490796	-0.229903	0.065460
0.072921	0.172246	-0.357751
0.104514	-0.463000	0.079367
-0.226080	-0.154659	-0.038422
0.406115	-0.192794	-0.441992
0.181755	0.088268	0.277574
-0.055334	0.491792	0.263102

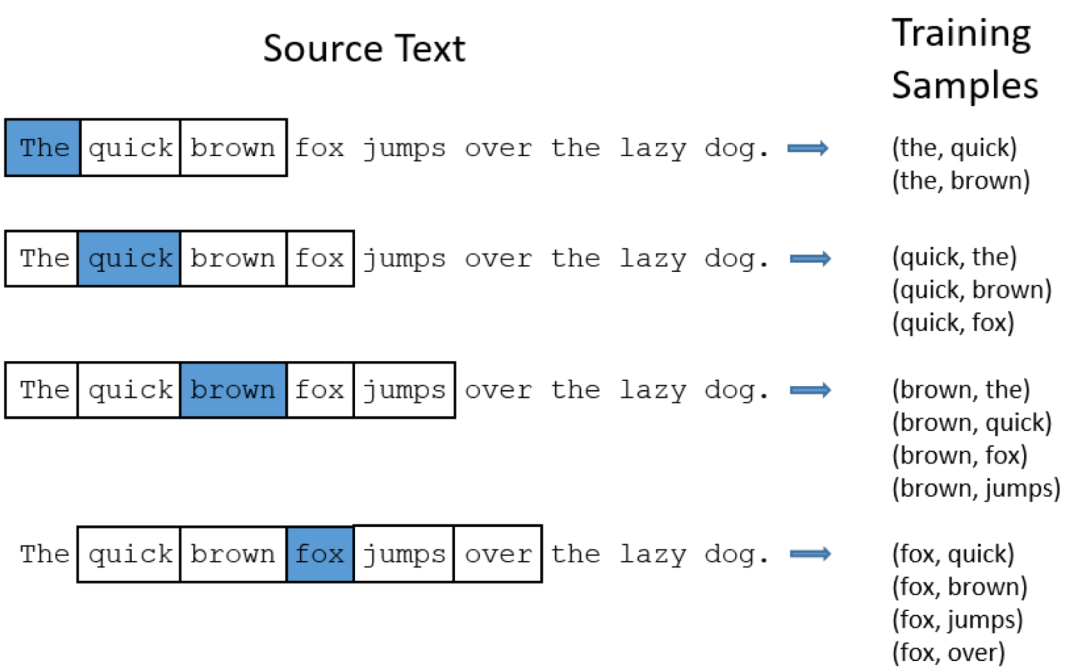
From imput to hidden layer: $\mathrm{H^T = X^T WI}$

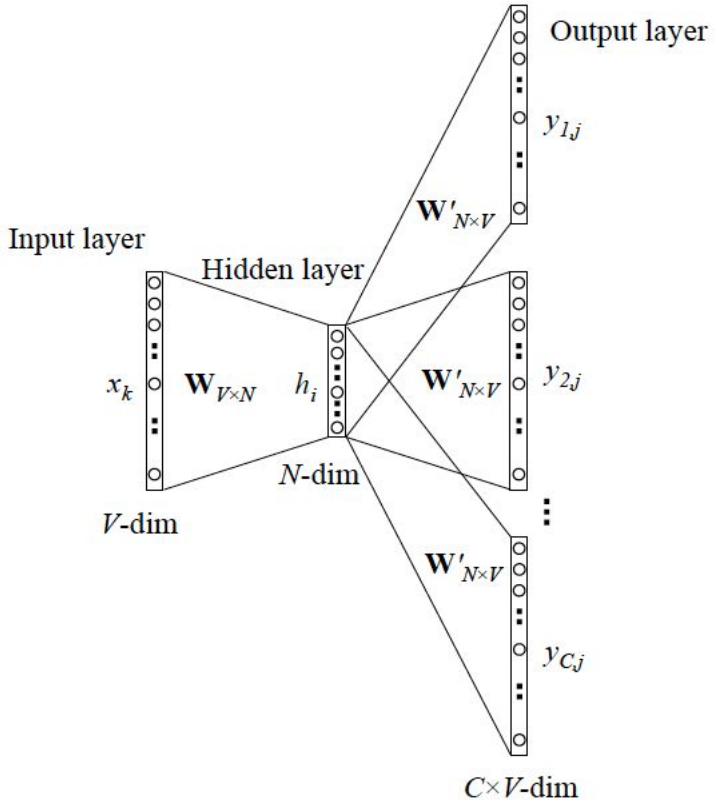
From hidden layer to output layer: $\mathrm{H^T WO}$

$$y_k = \Pr(word_k \mid word_{context}) = \frac{\exp(activation(k))}{\sum_{n=1}^{V} \exp(activation(n))}$$

PHILIPS

Word2Vec – Skip-gram Model





PHILIPS

Further Improvement enabled by Word2Vec

- Pairing the ingredients from two recipes, r_1 and r_2, in descending similarity order

r_2 \ r_1	gilthead filet	lemon juice	egg	oil	flour
white fish filet	0.83	0.37	0.29	0.2	0.29
tortilla chips	0.35	0.34	0.37	0.16	0.43
egg	0.29	0.32	1	0.19	0.39
oil	0.11	0.19	0.19	1	0.27
pepper	0.33	0.42	0.24	0.19	0.42

Remove ingredients that are already paired

r_2 \ r_1	gilthead filet	lemon juice	egg	oil	flour
white fish filet	0.83	0.37	0.29	0.2	0.29
tortilla chips	0.35	0.34	0.37	0.16	0.43
~~egg~~	~~0.29~~	~~0.32~~	~~1~~	~~0.19~~	~~0.39~~
oil	0.11	0.19	0.19	1	0.27
pepper	0.33	0.42	0.24	0.19	0.42

Repeat this process until ingredients in either of recipes are exhausted

r_2 \ r_1	gilthead filet	lemon juice	egg	oil	flour
white fish filet	0.83	0.37	0.29	0.2	0.29
tortilla chips	0.35	0.34	0.37	0.16	0.43
~~egg~~	~~0.29~~	~~0.32~~	~~1~~	~~0.19~~	~~0.39~~
~~oil~~	~~0.11~~	~~0.19~~	~~0.19~~	~~1~~	~~0.27~~
pepper	0.33	0.42	0.24	0.19	0.42

Paired ingredients

r_1	r_2	Similarity
egg	egg	1
oil	oil	1
gilthead filet	white fish filet	0.83
tortilla chips	flour	0.43
pepper	lemon juice	0.42

PHILIPS

Modified Similarity Indices

- Main ingredient similarity index between recipes r_1 and r_2:

$$\mathrm{Sim}(r_1, r_2) = \frac{\sum_{i,j \text{ are paired (main) ingredients}} \min(w_{r_1 i}, w_{r_2 j}) \cdot \log(P_i \cdot P_j) \cdot \mathrm{Sim}(i,j)}{\sum_{i,j \text{ are paired (main) ingredients}} \max(w_{r_1 i}, w_{r_2 j}) \cdot \log(P_i \cdot P_j) + 2\sum_{k \text{ is unpaired (main) ingredient in } r_1 \text{ or } r_2} \max(w_{r_1 k}, w_{r_2 k}) \cdot \log(P_k)}$$

- Flavor similarity index between recipes r_1 and r_2:

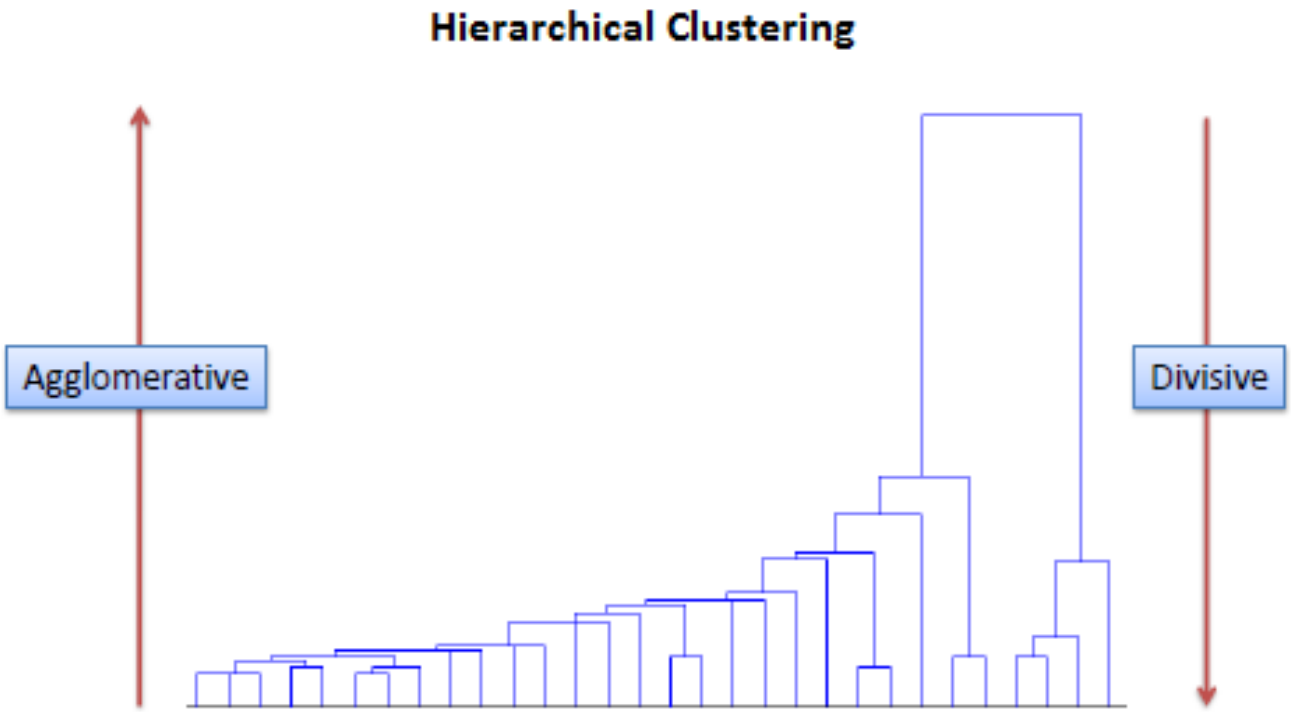
$$\mathrm{Sim}(r_1, r_2) = \frac{\sum_{i,j \text{ are paired flavors}} \min(w_{r_1 i}, w_{r_2 i}) \cdot \log\left((1-P_i)\cdot(1-P_j)\right) \cdot \mathrm{Sim}(i,j)}{\sum_{i,j \text{ are paired } flavors} \max(w_{r_1 i}, w_{r_2 i}) \cdot \log\left((1-P_i)\cdot(1-P_j)\right) + 2\sum_{k \text{ is unpaired flavors in } r_1 \text{ or } r_2} \max(w_{r_1 k}, w_{r_2 k}) \cdot \log(1-P_k)}$$

- If $\mathrm{Sim}(i,j) = 1$ for $i = j$ and $\mathrm{Sim}(i,j) = 0$ otherwise, the similarity indices will reduce to the previous situation

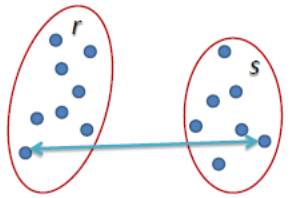
	Fish lasagna with beets,MR-931538C8	Honey & Lemon roasted chicken,MR-A750E68D	Spring Rolls,MR-2942CCA5	…
Fish lasagna with beets,MR-931538C8	1	0.108006999	0.12778	…
Honey & Lemon roasted chicken,MR-A750E68D	0.108006999	1	0.22112	…
Spring Rolls,MR-2942CCA5	0.127779549	0.221120079	1	…
…	…	…	…	…

PHILIPS

Hierarchical Clustering

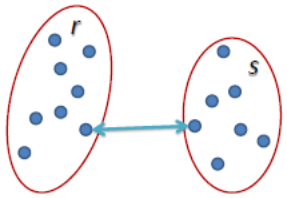


‘Complete’



$$L(r,s) = \max(D(x_{ri}, x_{sj}))$$

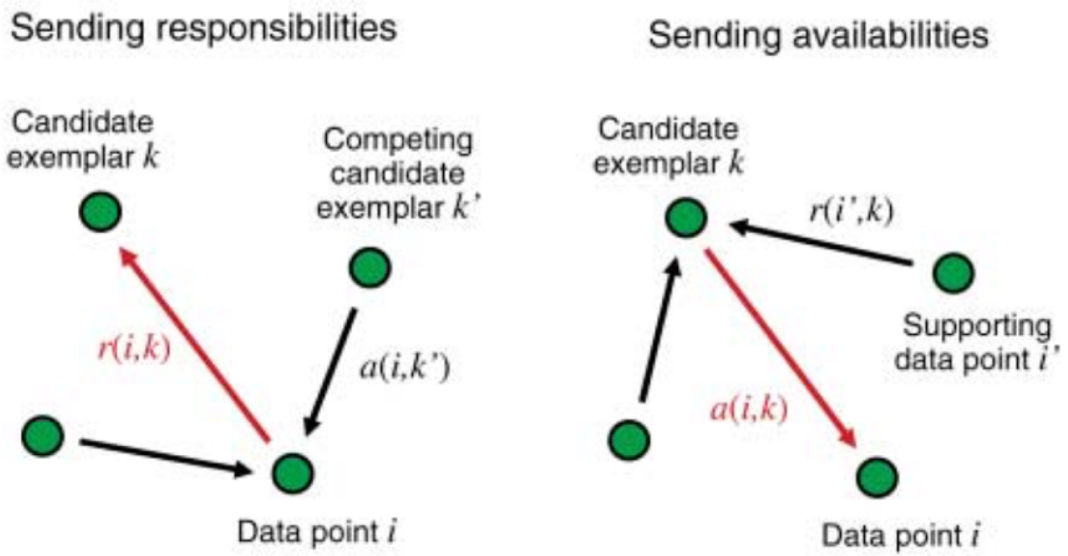
‘Single’



$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$

PHILIPS

Affinity Propagation



$$a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i' \text{ s.t. } i' \notin \{i,k\}} \max\{0, r(i',k)\}\right\}$$

$$r(i,k) \leftarrow s(i,k) - \max_{k' \text{ s.t. } k' \neq k} \{a(i,k') + s(i,k')\}$$

$$a(k,k) \leftarrow \sum_{i' \text{ s.t. } i' \neq k} \max\{0, r(i',k)\}$$

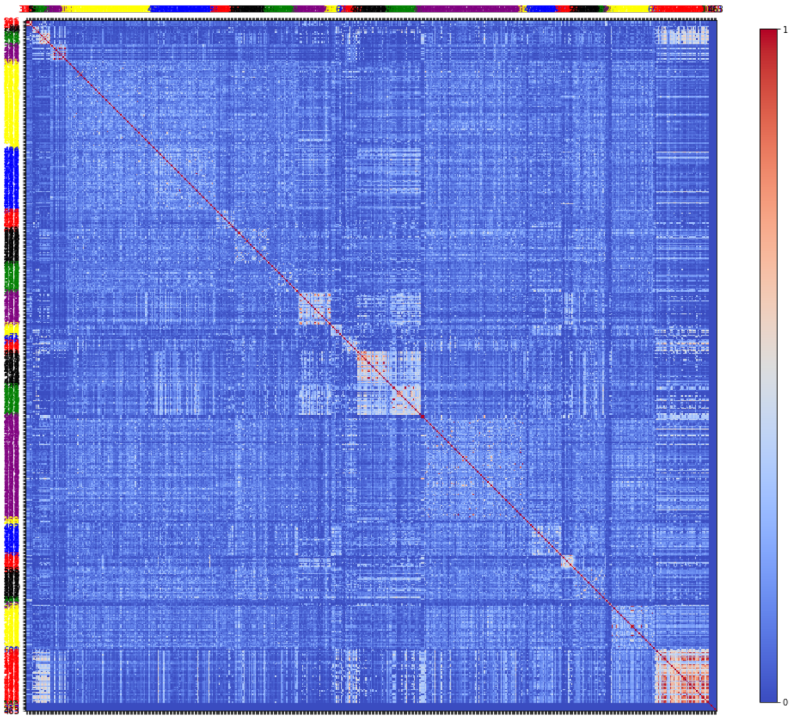
$$r_{t+1}(i,k) \leftarrow (1-\lambda) r_{t+1}(i,k) + \lambda r_t(i,k)$$

$$a_{t+1}(i,k) \leftarrow (1-\lambda) a_{t+1}(i,k) + \lambda a_t(i,k)$$

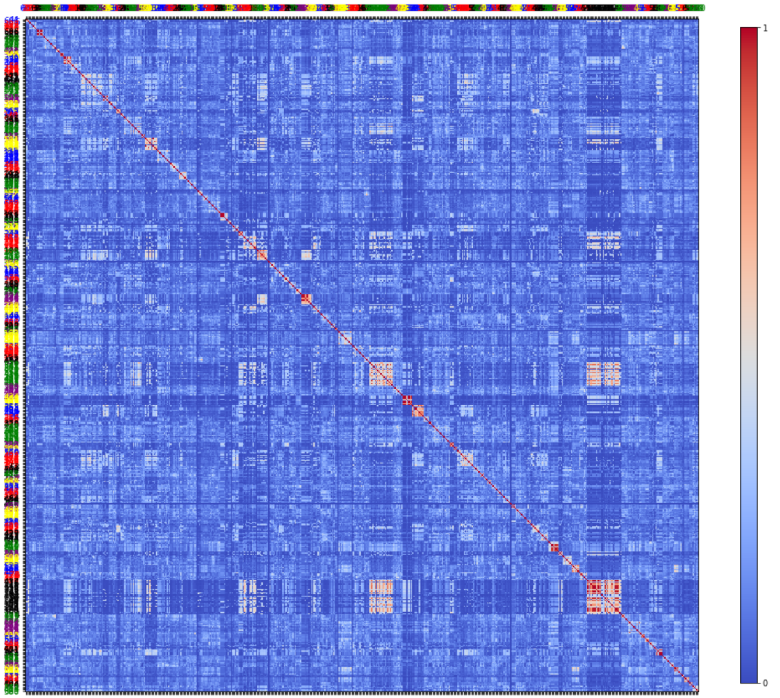
PHILIPS

Clustering Method Comparison

Hierarchical Clustering



Affinity Propagation



PHILIPS

Clustering Result (Affinity Propagation)

Honey & Lemon roasted chicken

Ingredients: 'lemon', 'apricot, dried', 'thyme, fresh', 'chicken', 'green courgette / zucchini', 'yellow courgette / zucchini', 'honey', 'salt', 'olive oil', 'apple', 'pepper', 'red onion', 'green garlic']

Main Ingredients: ‘chicken’

Seasonings: ‘salt’

Overall	Overall (Word2vec)	Main	Main (Word2vec)
Honey mustard chicken	Korean Fried Chicken	Honey mustard chicken	Korean Fried Chicken
Orange & paprika chicken	Chicken Tikka	Orange & paprika chicken	Honey mustard chicken
	Honey mustard chicken		Frittata with goat cheese tomatoes and pesto sauce
	Frittata with goat cheese tomatoes and pesto sauce		Korean BBQ Satay
	Korean BBQ Satay		Chicken Kofta
	Chicken Kofta		Green Curry Chicken
	Mini Peppers with Goat Cheese		Country Chicken Tenders
	Jerk Chicken Wings		Ricotta Balls
	Ricotta Balls		

Clustering Result (Affinity Propagation)

Ingredients: 'lemon', 'apricot, dried', 'thyme, fresh', 'chicken', 'green courgette / zucchini', 'yellow courgette / zucchini', 'honey', 'salt', 'olive oil', 'apple', 'pepper', 'red onion', 'green garlic']

Main Ingredients: ‘chicken’

Seasonings: ‘salt’

Flavor	Flavor (Word2vec)	Seasoning	Seasoning (Word2vec)
Chicken and bell pepper kebab with couscous	Honey mustard chicken	Baked belly pork slices with parsley potatoes	Baked belly pork slices with parsley potatoes
Grilled corn with lemon chicken	Grilled Caesar salad with chicken	Honey mustard chicken	Honey mustard chicken
Stuffed pumpkin with spinach and goat cheese	Grilled Caesar salad with chicken	Savoy Cabbage Roulade with Veal Filling	Savoy Cabbage Roulade with Veal Filling
Baba ganoush		Easter Bunny with egg	Winter salad with salmon and lentils
Salmon with fennel		Winter salad with salmon and lentils	Stuffed avocados with salmon and Mexican tostadas
Potato and chicken salad with spring vegetables		Stuffed avocados with salmon and Mexican tostadas	Meat kebabs with lecsó
		Meat kebabs with lecsó	Veal meatballs with Asian Pumpkin & vegetables
		Veal meatballs with Asian Pumpkin & vegetables	Lasagna with salmon and broccoli
		Lasagna with salmon and broccoli	Colorful Vegetable Strudel
		Colorful Vegetable Strudel	Spanish chicken with potato and parsnip crisps

Ranking Result

Honey & Lemon roasted chicken

Ingredients: 'lemon', 'apricot, dried', 'thyme, fresh', 'chicken', 'green courgette / zucchini', 'yellow courgette / zucchini', 'honey', 'salt', 'olive oil', 'apple', 'pepper', 'red onion', 'green garlic']

Main Ingredients: ‘chicken’

Seasonings: ‘salt’

Overall	Overall (word2vec)	Main	Main (word2vec)
Honey mustard chicken'	Honey mustard chicken'	Honey mustard chicken'	Honey mustard chicken'
'Orange & paprika chicken '	'Chicken drumettes'	'Orange & paprika chicken '	'Orange & paprika chicken '
'Rice Pudding with Apple'	'Orange & paprika chicken '	'Fish lasagna with beets'	'Chicken drumsticks in a honey/orange marinade '
'Verdure al forno with anchovy'	'Chicken with Riesling and elderberry'	'Roast Beef with vegetables'	'Chicken drumettes'
'Muesli bars'	'Fried Spicy Thai Chicken Wings'	'Baked Dumplings'	'Chicken with Riesling and elderberry'
'Lemon Muffin Cakes '	'Chicken drumsticks in a honey/orange marinade '	'Chocolate Mug Cakes'	'Fried Spicy Thai Chicken Wings'
'Lemon Muffin Cakes'	'Korean Fried Chicken'	'Bacon Wrapped Prawns'	'Curry chicken skewers'
'Airfryer Fish tikka masala'	'Spicy Wings'	'Pumpkin oven cheese gratin'	'Korean Fried Chicken'
'Norwegian Apple Pie'	'Spicy drumsticks'	'Breaded button mushrooms with rocket and tomato salad'	'Chicken Nuggets'

Ranking Result

Ingredients: 'lemon', 'apricot, dried', 'thyme, fresh', 'chicken', 'green courgette / zucchini', 'yellow courgette / zucchini', 'honey', 'salt', 'olive oil', 'apple', 'pepper', 'red onion', 'green garlic']

Main Ingredients: ‘chicken’

Seasonings: ‘salt’

Flavor	Flavor (word2vec)	Seasoning	Seasoning (word2vec)
Honey mustard chicken'	Honey mustard chicken'	Spanish chicken with potato and parsnip crisps'	Spanish chicken with potato and parsnip crisps'
'Chicken with Riesling and elderberry'	'Potato and chicken salad with spring vegetables'	'Savoy Cabbage Roulade with Veal Filling'	'Savoy Cabbage Roulade with Veal Filling'
'Chicken drumsticks in a honey/orange marinade '	'Stuffed pumpkin with spinach and goat cheese'	'Colorful Vegetable Strudel'	'Colorful Vegetable Strudel'
'Chicken and bell pepper kebab with couscous'	'Grilled Caesar salad with chicken'	'Honey mustard chicken'	'Honey mustard chicken'
'Baba ganoush'	'Grilled Caesar salad with chicken'	'Winter salad with salmon and lentils'	'Winter salad with salmon and lentils'
'Fried Pineapple'	'Stuffed Portobello with roasted radicchio'	'Baked belly pork slices with parsley potatoes '	'Baked belly pork slices with parsley potatoes '
'Steak shoarma with paprika sauce'	'Chicken with Riesling and elderberry'	'Lasagna with salmon and broccoli'	'Lasagna with salmon and broccoli'
'Salmon with fennel'	'Carrots glazed with thyme honey'	'Quiche Lorraine'	'Quiche Lorraine'
'Roasted winter vegetables'	'Salmon with fennel'	'Meat kebabs with lecsó'	'Meat kebabs with lecsó'

References

http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/
https://iksinc.online/tag/word2vec/
https://nlp.stanford.edu/IR-book/html/htmledition/hierarchical-clustering-1.html
https://www.saedsayad.com/clustering_hierarchical.htm
Brendan J. Frey and Delbert Dueck, “Clustering by Passing Messages Between Data Points”, Science Feb. 2007

Special thanks to my mentor!

Thank you!


PHILIPS