

Doing Research in Natural Language Processing

Session 14: Scientific Writing: Graphs, Diagrams, Tables

Frank Keller

3 November 2023

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Graphs

Diagrams

Captions

Tables

Reading: Zobel (2014), ch. 11.

Please also look at Alley's web site, which has a lot of videos and additional materials:

<https://www.craftofscientificwriting.com/>

Visual material is an important part of a paper:

- diagrams illustrate complex ideas, processes, or models
- graphs show trends or relationships in data
- tables present results or regularities in data
- textual panels present algorithms or mathematical formulas

Such materials attracts the reader's attention; some readers will only look at figures and tables, they will not read (all of) the text.

Graphs

Graphs

- graphs can make behaviors and trends obvious that a hard to discern from a table
- keep graphs simple, avoid both clutter and unnecessary whitespace
- for elements such as secondary ticks, legends, gridlines, boxes, ask if you really need them
- use the same fonts in graphs and tables as in the main text
- sometimes logarithmic axes are appropriate
- a table of results can often be represented as a bar graph
- if you use multiple graphs to display the same quantity, use the same axis (same range) in all of them

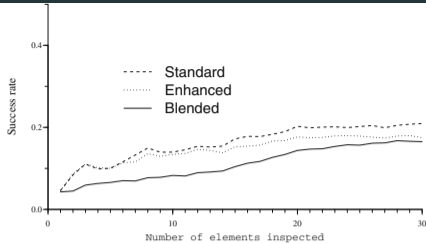


FIGURE 7. Success rate as the number of inspected items is increased. It is clear that blending is not effective.

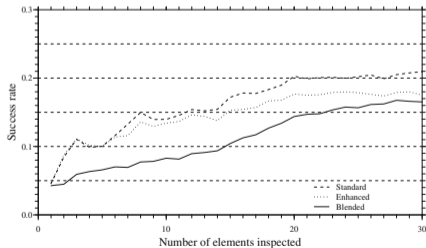


FIGURE 7. Success rate as the number of inspected items is increased. It is clear that blending is not effective.

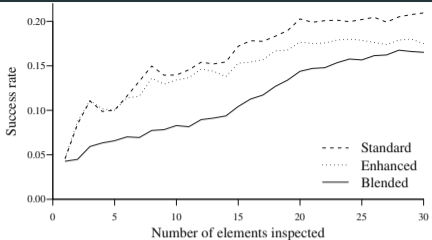


FIGURE 7. Success rate as the number of inspected items is increased. It is clear that blending is not effective.

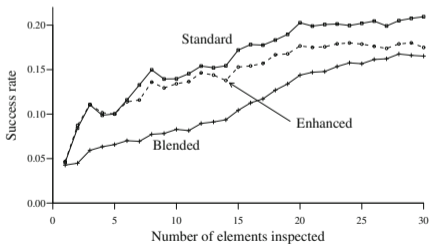


FIGURE 7. Success rate as the number of inspected items is increased. It is clear that blending is not effective.

✗

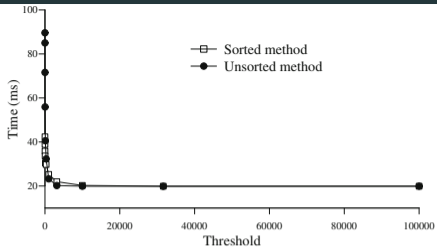


FIGURE 6. Evaluation time (in milliseconds) for bulk insertion methods as threshold is varied.

✓

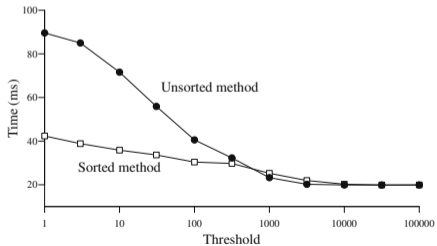


FIGURE 6. Evaluation time (in milliseconds) for bulk insertion methods as threshold is varied.

Data set	Method	
	A	B
Small, random	11.5	11.6
Large, random	27.9	17.1
Small, clustered	9.7	8.2
Large, clustered	24.0	13.5
All documents	49.4	60.1
First 1000	21.1	35.4
Last 1000	1.0	5.5

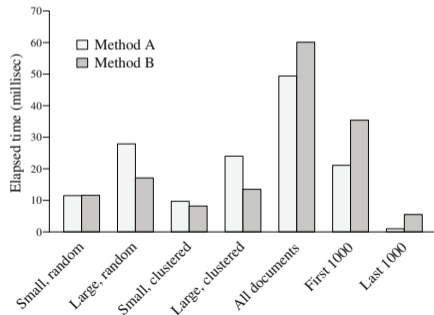


FIGURE 2. Elapsed time (milliseconds) for methods A and B applied to data sets 1–7.

Diagrams

Diagrams

- diagrams show architectures, structures, processes, relationships, or states
- typically, the diagram should just show **one** of the things; an attempt to combine them often makes the diagram less clear
- it's a good idea to sketch the diagram by hand first, check layout, proportions, use of space, sizes of elements
- focus on the concept being illustrated, avoid clutter and unnecessary detail
- use pictorial elements consistently (arrows or boxes of the same kind always have the same meaning, etc.)
- don't expect to get it right first time, revise your diagrams as you would revise your text

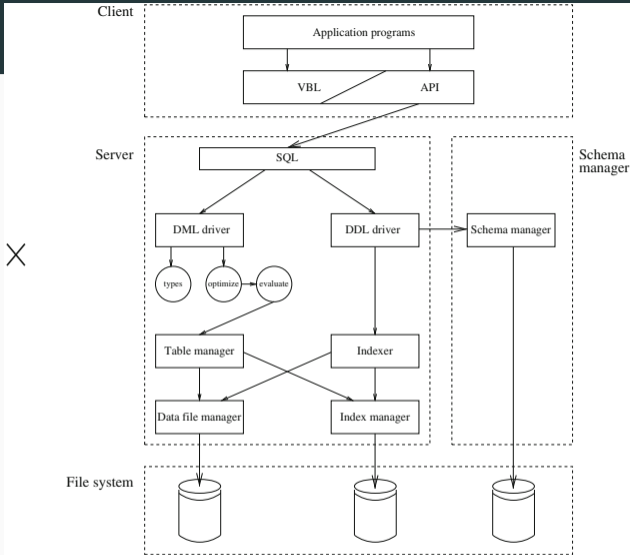


FIGURE 1.3. System architecture, showing the relationship between the major components. Each component is an independent process. Note the lack of a single interface to the file system.

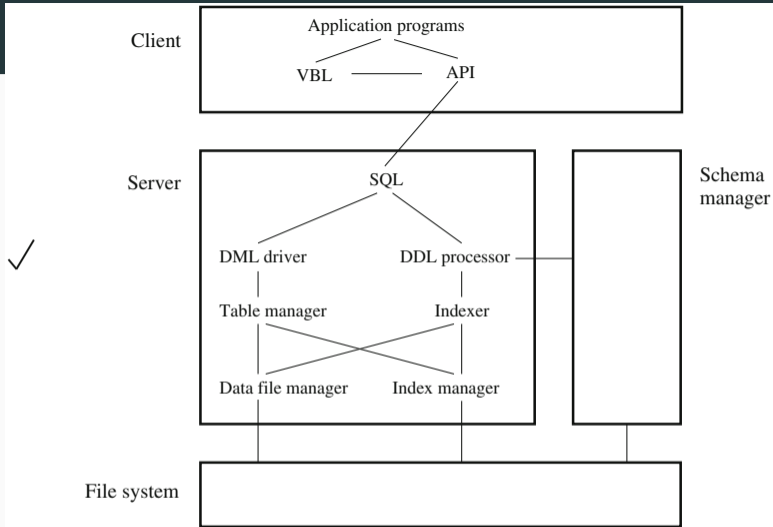


FIGURE 1.3. System architecture, showing the relationship between the major components. Each component is an independent process. Note the lack of a single interface to the file system.

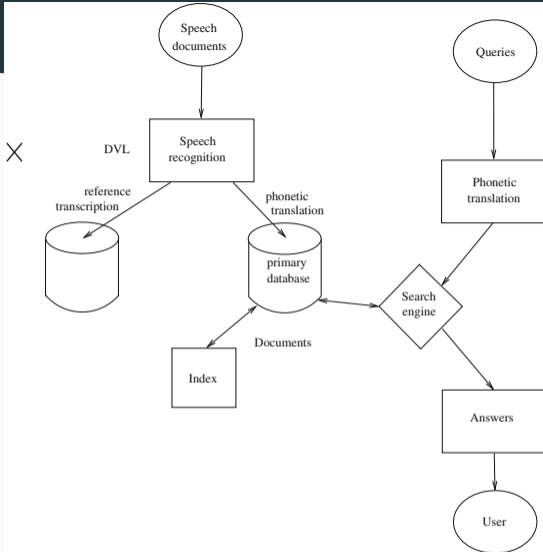


FIGURE 7. The QUIRK system for matching written queries to speech. Each input document is translated into a string of phonemes and then stored. Queries are also translated into phonemes, which can be matched to the documents. Answers are returned to the user.

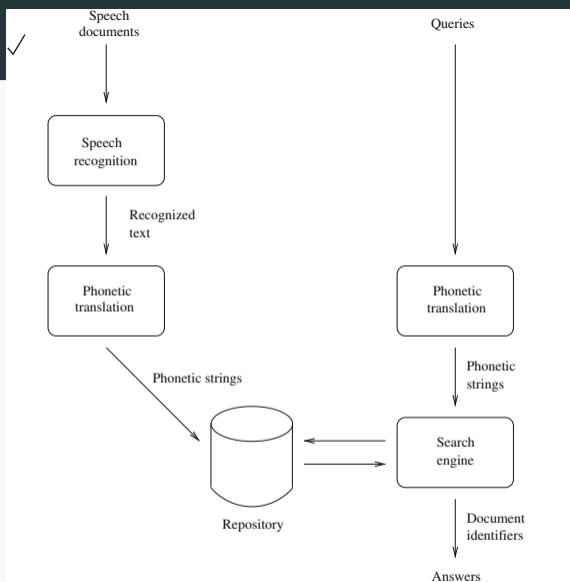


FIGURE 7. The QUIRK system for matching written queries to speech. Each input document is translated into a string of phonemes and then stored. Queries are also translated into phonemes, which can be matched to the documents.

Captions

Captions

- captions should fully describe the major elements of a figure or table
- together with its caption, the figure or table should be self-contained, i.e., understandable without referring to the text
- captions should assist a reader who's only skimming the paper, or who is going back to re-read parts of a longer paper
- normally, the caption appears above a table, but below a figure
- if you use abbreviations or symbols in a figure or table, then these need to be explained in the caption
- the caption can also contain additional detail that would interrupt the flow of the main text

Over to You

Exercise 1

Remember the diagram from Vaswani et al. (2017) (next page):

- Is this a well-designed diagram?
- Does it have the right level of complexity?
- Is the caption appropriate?

How would you modify the diagram and caption to improve it?

Exercise 1

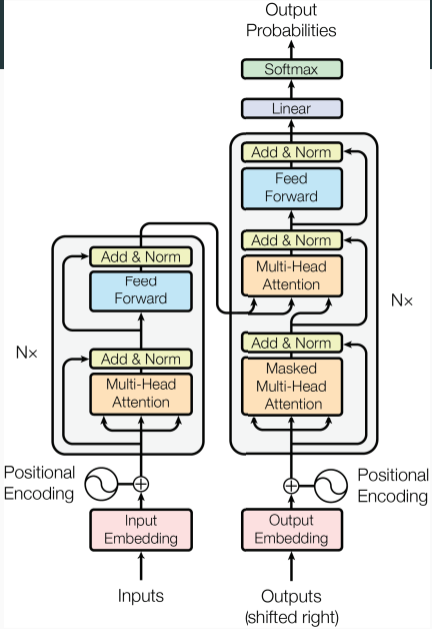


Figure 1: The Transformer - model architecture.

Tables

Tables

- some information cannot be presented easily in graphs or diagrams
- in some cases, the exact numeric values are important
- tables are more suitable than graphs if only a small number of values need to be displayed
- tables can have a hierarchical structure: columns and rows can be partitioned or have internal structure
- the structure needs to be indicated by headings, labels, dividers
- limit the use of horizontal rules; vertical rules should be avoided; tables should contain sufficient whitespace
- don't make a table too big; instead, use two tables or a graph

TABLE 6. *Statistics of text collections used in experiments.*

X

STATISTICS	SMALL	LARGE
Characters	18,621	1,231,109
Words	2,060	173,145
After stopping	1,200	98,234
Index size	1.31 Kb	109.0 Kb

TABLE 6. *Statistics of text collections used in experiments.*

✓

	Collection	
	Small	Large
File size (Kb)	18.2	1,202.3
Index size (Kb)	1.3	109.0
Number of words	2,060	173,145
— after stopping	1,200	98,234

TABLE 2.1. *Impact on performance (processing time and effectiveness) of varying each of the three parameters in turn, for both data sets. Default parameter values are shown in parentheses. Note that $p = 100,000$ is not meaningful for the data set SINGLE.*

Parameter	Data set			
	SINGLE		MULTIPLE	
	CPU (msec)	Effective (%)	CPU (ms)	Effective (%)
n ($k = 10, p = 100$)				
2	57.5	55.5	174.2	22.2
3	21.5	50.4	79.4	19.9
4	16.9	47.5	66.1	16.3
k ($n = 2, p = 100$)				
10	57.5	55.5	174.2	22.2
100	60.0	56.1	163.1	21.3
1000	111.3	55.9	228.8	21.4
p ($n = 2, k = 10$)				
100	57.5	55.5	174.2	22.2
1000	13.8	12.6	19.8	2.1
10,000	84.5	56.0	126.4	6.3
100,000	—	—	290.7	21.9

✓

TABLE 11. *Resources used during compression and indexing. Only the vocabulary is constructed in the first pass; the other structures are built in the second pass.*

Pass	Output	Size		CPU Hr:Min	Mem Mb
		Mb	%		
Pass 1:					
Compression	Model	4.2	0.2	2:37	25.6
Inversion	Vocabulary	6.4	0.3	3:02	18.7
Overhead				0:19	2.5
Total		10.6	0.5	5:58	46.8
Pass 2:					
Compression	Text	605.1	29.4	3:27	25.6
	Doc. map	2.8	0.1		
Inversion	Index	132.2	6.4	5:25	162.1
	Index map	2.1	0.1		
	Doc. lens	2.8	0.1		
	Approx. lens	0.7	0.0		
Overhead				0:23	2.5
Total		745.8	36.3	9:15	190.2
Overall		756.4	36.8	15:13	190.2

X

TABLE 11. *Resources used during compression and indexing. Only the vocabulary is constructed in the first pass; the other structures are built in the second pass.*

Task	Size (Mb)	CPU (Hr:Min)	Memory (Mb)
<i>Pass 1:</i>			
Compression	4.2	2:37	25.6
Inversion	6.4	3:02	18.7
Overhead	—	0:19	2.5
Total	10.6	5:58	46.8
<i>Pass 2:</i>			
Compression	607.9	3:27	25.6
Inversion	137.8	5:25	162.1
Overhead	—	0:23	2.5
Total	745.8	9:15	190.2
Overall	756.4	15:13	190.2

Over to You

Exercise 2

Here are some tables from Vaswani et al. (2017) (next pages):

- Is the table layout good? How about the use of whitespace?
- Can you decode the hierarchical structure of these tables?
- Should they maybe have used a graph instead?
- Are the captions appropriate?

How would you modify the tables and captions to improve them?

Exercise 2

Table 2: The Transformer achieves better BLEU scores than previous state-of-the-art models on the English-to-German and English-to-French newstest2014 tests at a fraction of the training cost.

Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	41.29	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$3.3 \cdot 10^{18}$	
Transformer (big)	28.4	41.8	$2.3 \cdot 10^{19}$	

Exercise 2

Table 3: Variations on the Transformer architecture. Unlisted values are identical to those of the base model. All metrics are on the English-to-German translation development set, newstest2013. Listed perplexities are per-wordpiece, according to our byte-pair encoding, and should not be compared to per-word perplexities.

	N	d_{model}	d_{ff}	h	d_k	d_v	P_{drop}	ϵ_{ls}	train steps	PPL (dev)	BLEU (dev)	params $\times 10^6$
base	6	512	2048	8	64	64	0.1	0.1	100K	4.92	25.8	65
(A)				1	512	512				5.29	24.9	
				4	128	128				5.00	25.5	
				16	32	32				4.91	25.8	
				32	16	16				5.01	25.4	
(B)					16					5.16	25.1	58
					32					5.01	25.4	60
(C)	2									6.11	23.7	36
	4									5.19	25.3	50
	8									4.88	25.5	80
		256			32	32				5.75	24.5	28
		1024			128	128				4.66	26.0	168
			1024							5.12	25.4	53
			4096						4.75	26.2	90	
(D)							0.0			5.77	24.6	
							0.2			4.95	25.5	
								0.0		4.67	25.3	
								0.2		5.47	25.7	
(E)				positional embedding instead of sinusoids						4.92	25.7	
big	6	1024	4096	16			0.3		300K	4.33	26.4	213

References

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*. Curran Associates, Red Hook, NY, pages 5998–6008.
- Zobel, Justin. 2014. *Writing for Computer Science*. Springer, New York, NY, 3 edition.