Natural Language Processing

NLP in one day

Introduction

Angus Roberts
Senior Lecturer,
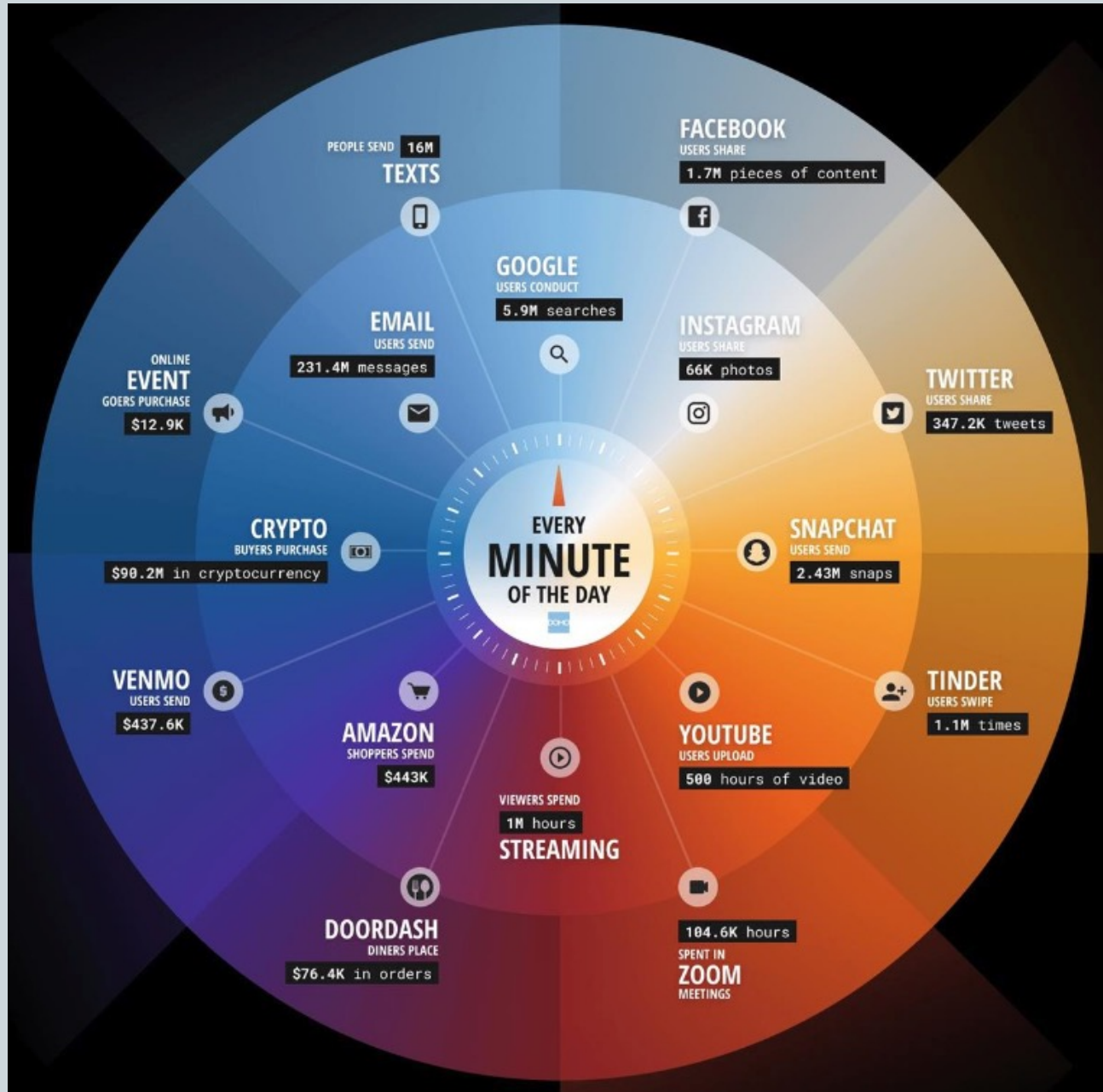Health Informatics

Yamiko Msosa
Research Fellow,
Health Informatics

Leo Xinyue Zhang
PhD Student,
Clinical NLP

Text is everywhere

Data Never Sleeps 10.0 from domo.com (April 2022)

Generative AI

Write a short definition of natural language processing

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language. It involves analyzing and processing natural language data, such as text or speech, to perform tasks like translation, sentiment analysis, speech recognition, and information extraction. NLP combines linguistics, computer science, and machine learning to bridge the gap between human communication and computer understanding.

Electronic health records (EHRs) – structured and unstructured information

HAEMATOLOGY
HAEMOGLOBIN (g/L)   151
HCT   0.422
RED CELL COUNT   4.83
MCV   87.4
MCH   31.3
MCHC (g/L)   ▬ * 358
RDW
PLATELET COUNT
MPV
WHITE CELL COUNT
  Neutrophils
  Lymphocytes
  Monocytes
  Eosinophils
  Basophils
ESR

BIOCHEMISTRY
SODIUM
POTASSIUM
CHLORIDE
BICARBONATE
UREA
CREATININE
estimated GFR

Clinical Findings

Concept ID	Preferred term
16932000	Nausea and vomiting
68566005	Urinary tract infectious disease
38341003	Hypertensive disorder
49436004	Atrial fibrillation
49218002	Hip pain
301011002	Escherichia coli urinary tract infect
40835002	Coffee ground vomiting
167667006	Fecal occult blood: negative

Procedures

Concept ID	Preferred term
52734007	Total replacement of hip
117010004	Urine culture
76009000	Esophagogastroduodenoscopy
91251008	Physical therapy procedure

```
Reason: CHECK ETT TUBE PLACEMENT, ?PNA, CHF
[**Signature 1**]
UNDERLYING MEDICAL CONDITION:
 85 y/o male s/p acute mi and catherization now in ccu with
cardiogenic shock
REASON FOR THIS EXAMINATION:
CHECK ETT TUBE PLACEMENT
?PNA
CHF
[**Signature 1**]
 FINAL REPORT
CLINICAL INDICATION:  Assess endotracheal tube placement in
patient with congestive heart failure.

Comparison is made to previous study of one day earlier.
An endotracheal tube is present, in satisfactory position.
A  Swan-Ganz  catheter  terminates  in  the  proximal  left
pulmonary artery and has been withdrawn in the  interval.
An intraaortic balloon pump terminates about 3.3 cm below
the superior aspect of the aortic knob, and a nasogastric
tube  terminates  in  the  region  of  the  gastroduodenal
junction.
```

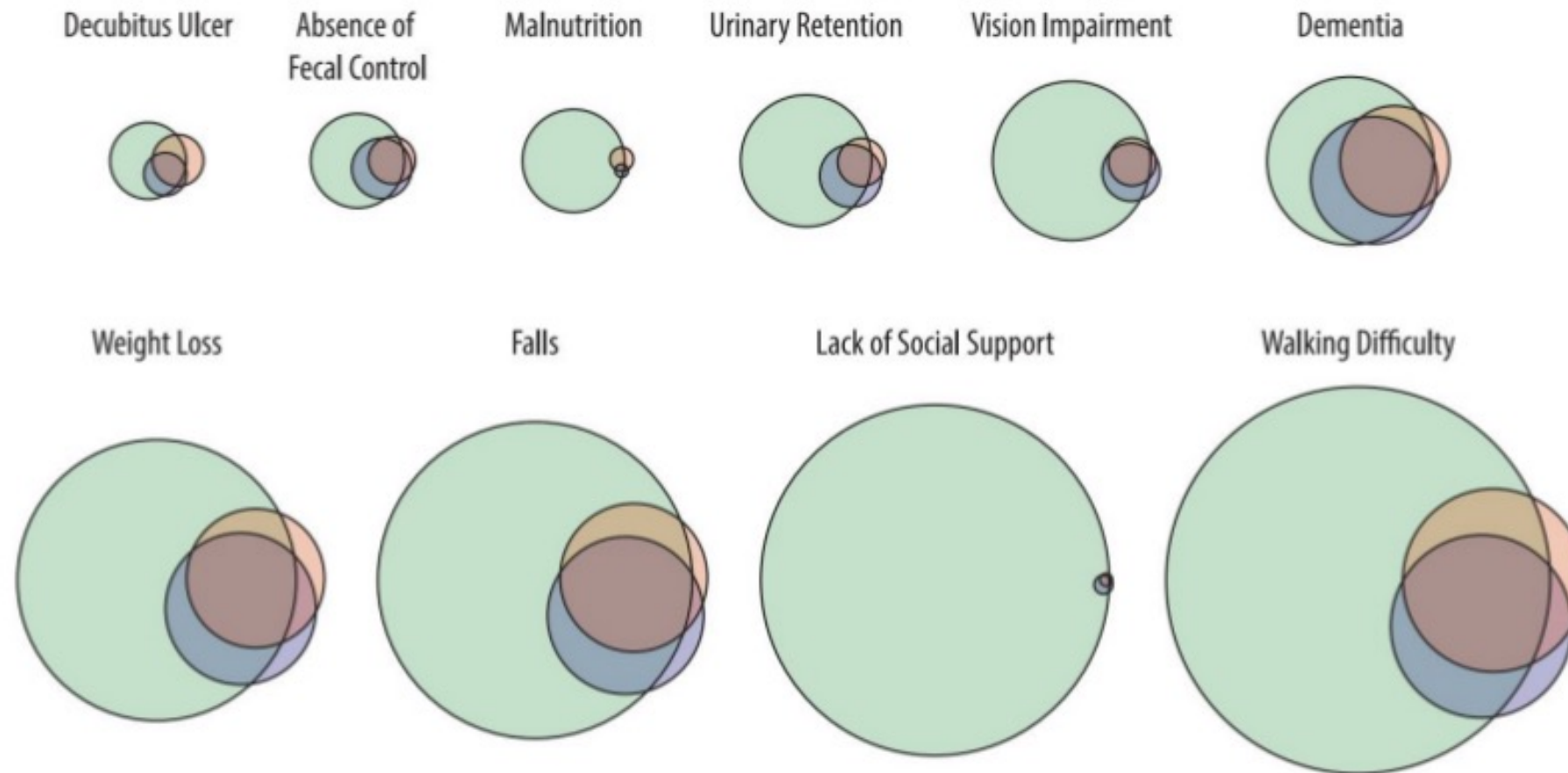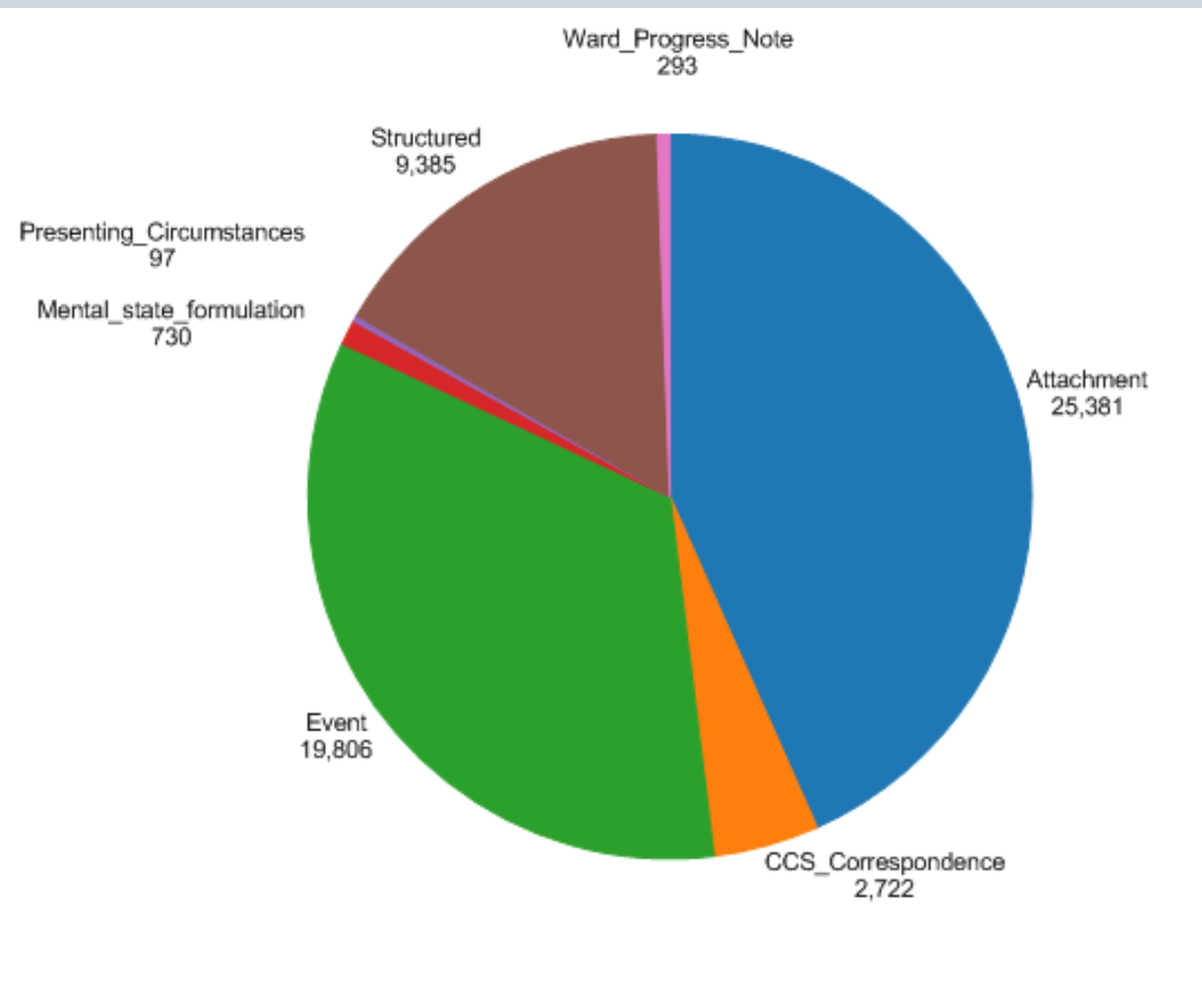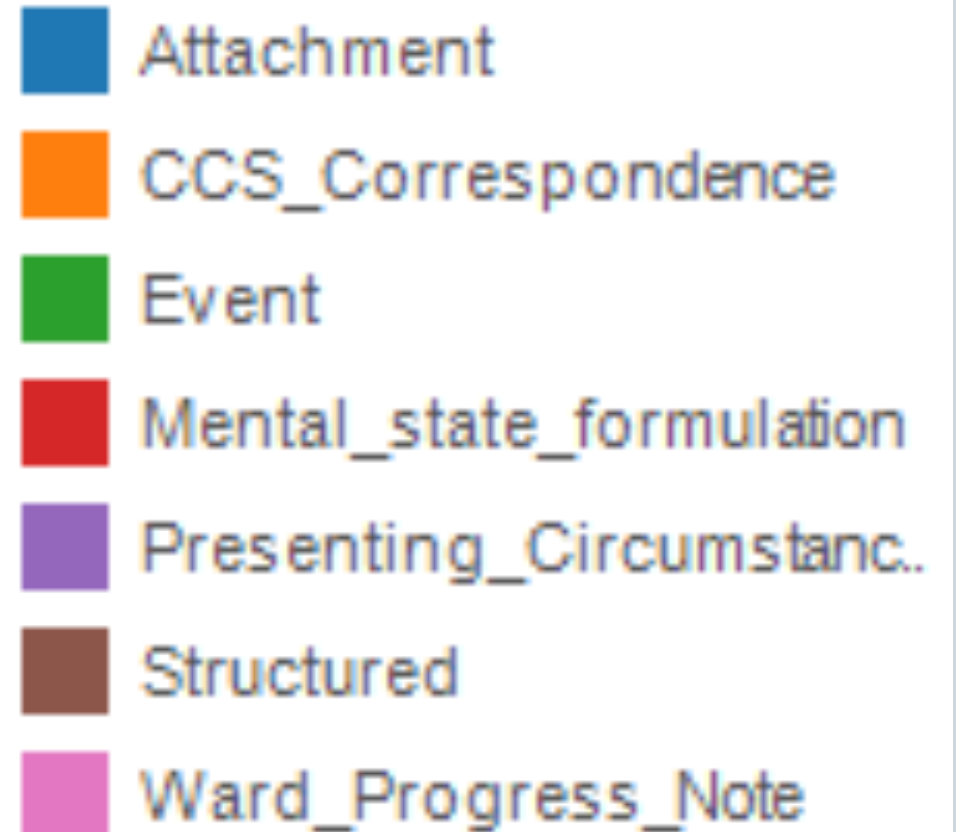# How much unstructured, textual information is in an EHR?

Figure 2. Green: Unstructured free text EHR data; Other colours: structured data. "The value of Unstructured Electronic Health Record Data in Geriatric Syndrome Case Identification". (Kharrazi et al., 2018)

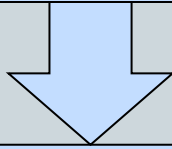# An example – Mini Mental State Examinations

This course

Pre-workshop - video lectures and study material
You did do all of it, didn't you?

This course

Pre-workshop - video lectures and study material
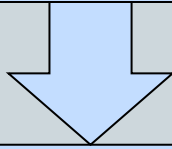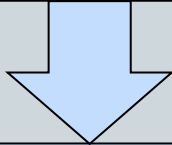*You did do all of it, didn't you?*

↓

Workshop – practicals
And a few linking presentations

This course

Pre-workshop - video lectures and study material
You did do all of it, didn't you?

⬇

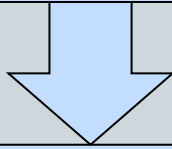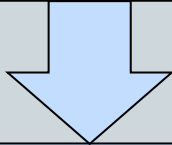Workshop – practicals
And a few linking presentations

⬇

Post-workshop – self-study reading material

This course

Pre-workshop - video lectures and study material
You did do all of it, didn't you?

⬇

Workshop – practicals
And a few linking presentations

⬇

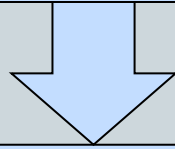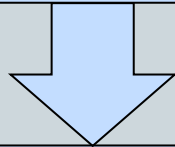Post-workshop – self-study reading material

Theory, methods

This course

Pre-workshop - video lectures and study material
You did do all of it, didn't you?

↓

Workshop – practicals
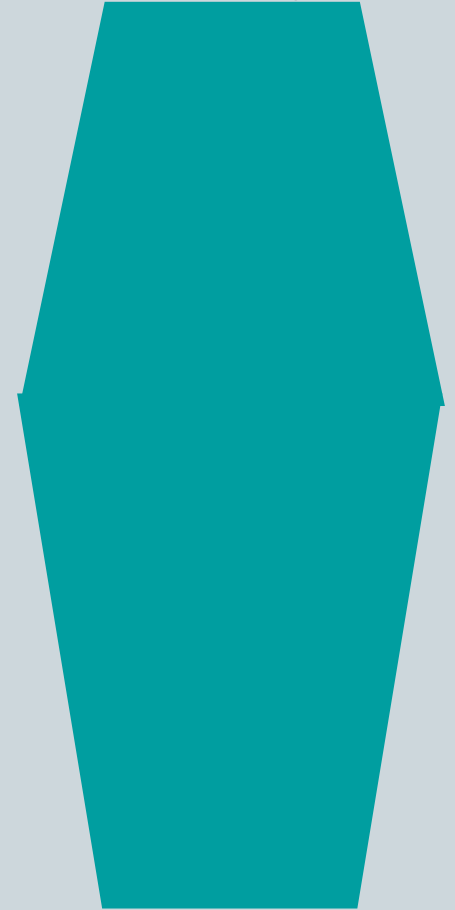*And a few linking presentations*

↓

Post-workshop – self-study reading material

Theory, methods

Practical techniques

Course material

All course material can be found on GitHub

Let's take a look:

https://github.com/KCL-Health-NLP/nlp-one-day-workshop/tree/main

Thank you

angus.roberts@kcl.ac.uk

https://www.kcl.ac.uk/people/angus-roberts