



Overview of CLEF eHealth 2012–2020 +2021

Assoc/Prof Hanna Suominen, Adj/Prof, PhD, MSc, SFHEA
Research Program & Team Leader in Machine Learning
hanna.suominen@anu.edu.au, @hajasu





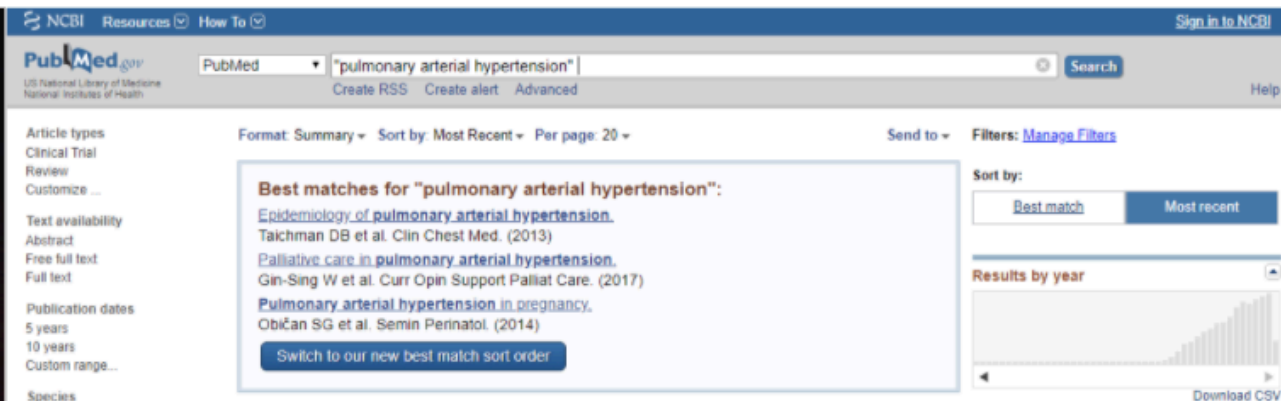
Martin Ollman



Martin Ollman



? Pulmonary arterial hypertension



Search

“Pulmonary arterial hypertension”
from 30+ million publications
(every day 2,200+ new publications)

returns

12,158 publications of which
2,755 are reviews and

935 reviews published within the last 5 yrs



Annual CLEF eHealth labs since 2012



?

(Lay)people's increasing difficulties to retrieve and digest valid & relevant information in their preferred language

?



From:

80 y/o male with 2 yr h/o SOB and GERD. Had Echo that revealed AS. Has had serial echo's since, with the last done on [. . .] which showed worsening AS.

To:

80 year old (y/o) male with 2-year (yr) history of (h/o) **shortness of breath**²⁶⁷⁰³⁶⁰⁰⁷ (SOB, [definition](#), [treatment guidelines](#)) and **heartburn**^{C0018834} (also known as **gastroesophageal reflux disease** (GERD, [definition](#), [treatment guidelines](#))).

Had a **diagnostic ultrasound of heart (echocardiogram)**¹⁶⁹²⁴¹⁰⁰⁰ = Echocardiogram abnormal)^{C0810219} (echo/Echo) that revealed **aortic valve stenosis**⁶⁰⁵⁷³⁰⁰⁴ (AS, [definition](#), [treatment guidelines](#), [guidelines on separating the symptoms of heart conditions from heartburn](#)). Has had serial **diagnostic ultrasound of heart (echocardiogram)**¹⁶⁹²⁴¹⁰⁰⁰)^{C0810219}'s since, with the last done on [. . .] which showed worsening **aortic valve stenosis**⁶⁰⁵⁷³⁰⁰⁴.



Timeline of the CLEF eHealth tasks



	2013	2014	2015	2016	2017	2018	2019	2020
IE	NER clinical reports	IE from clinical reports	NER biomed articles		Multilingual IE from health reports			
	Acronyms' normalization			Classification of death reports				
				Replication task				
Information Management		Health data visualisation	Nurses' handover reports management					
IR	Patient-centered IR/ Consumer Health Search							
		Cross-Lingual (Multilingual) IR						
					Technologically Assisted Reviews in Empirical Medicine			

Registered Teams

170

220

100

116

67

70

67

57

Submitting Teams

53

24

20

20

32

28

9

55



Information extraction & management



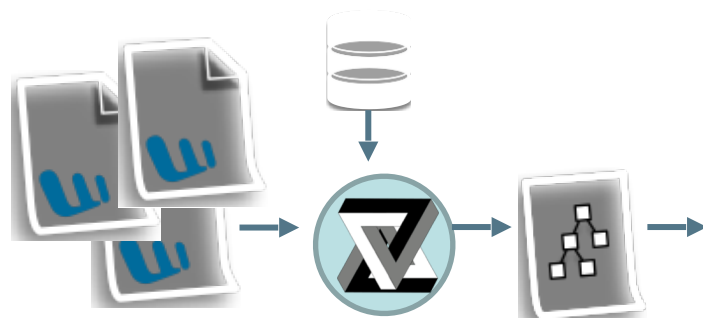
Improved usefulness of eHealth records through clinical coding

From clinical text

Through
methods &
models

To your
customised insight:

- Structuring → Data mining
- Summarisation
- Situational awareness
- Discovery
- Decision support





Why would I take CLEF eHealth if I do not study health applications?



3 FACTORS LEAD TO BETTER
PERFORMANCE & PERSONAL
SATISFACTION...

AUTONOMY
MASTERY
PURPOSE



- Large **societal impact**: advances in these tasks could contribute to health, healthcare, & societal prosperity
- Show that **my techniques work across** different domains and data/tasks, including health
 - generalisability
 - research funding
- Often our tasks are excellent instances of **fundamental problems** in information management, IE, and IR
 - authentic, timely, important, complex, ...
 - natural language processing, machine learning, evaluation, trust, cognitive biases



<https://youtu.be/u6XAPnuFjJc>



CLEF eHealth 2020 tasks



Co-chairs: Lorraine Goeuriot, Hanna Suominen, & Liadh Kelly

Task 1: **Multilingual Information Extraction (IE)**



International Classification of Diseases,
Clinical Modification (ICD-CM) coding of
clinical documents in Spanish

Co-Leaders: Martin Krallinger & Antonio Miranda

Task 2: **Consumer Health Search (IR)**



Spoken, speech-recognised, & typed queries
in a rich range of native English accents

Information topicality, readability, & credibility

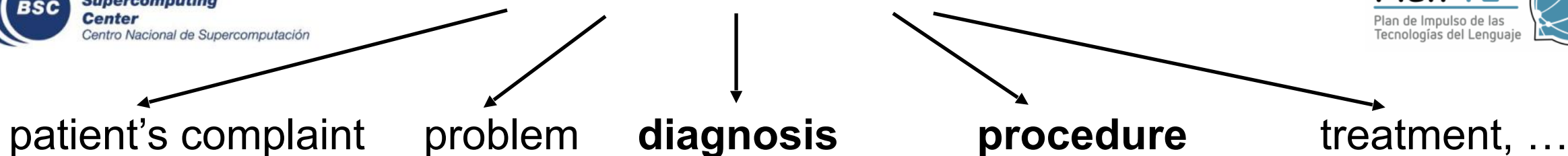
Co-Leaders: Lorraine Goeuriot & Hanna Suominen





Task 1 (IE) on CodiESP corpus

Transforming clinical text into a structured format using internationally recognised concepts or class codes



1,000 = 500 + 250 + 250

clinical case studies

chosen & codified by a clinical expert

- **Diagnosis** codes belong to ICD10-CM/CIE10 Diagnóstico
- **Procedure** codes belong to ICD10/CIE10 Procedimiento (ICD10-PCS) codes

18,483 (3,427 unique) annotated codes

1	Recién nacida a término que fue enviada para valoración oftalmológica a los 20 días de vida por la aparición de DIAGNOSTICO nistagmo ocasional.
2	Su madre, sin antecedentes de importancia durante el embarazo, había sido operada de DIAGNOSTICO glaucoma a los 20 años de edad con trabeculectomía bilateral y cirugía reconstructiva de DIAGNOSTICO sindactilia de manos.
3	En el examen oftalmológico se observaban microcórneas con aspecto discretamente velado, DIAGNOSTICO nistagmo horizontal ocasional, tono ocular digital PROCEDIMIENTO aumentado en forma bilateral y DIAGNOSTICO fondo de ojo (FO) con papilas de características conservadas; y el examen físico general revelaba una pirámide nasal estrecha con narinas estrechas y alas nasales finas, DIAGNOSTICO micrognatia leve y DIAGNOSTICO sindactilia de los dedos cuarto y quinto de ambas manos. PROCEDIMIENTO
4	Bajo sedación, la tonometría puso de manifiesto una presión intraocular (PIO) de 35 y 40mm Hg en los ojos DIAGNOSTICO derecho e izquierdo, respectivamente, y además se verificaron DIAGNOSTICO edema epitelial y nubéculas corneales superficiales.
6	Se realizó una trabeculectomía bilateral con buen control post-operatorio de la PIO, sin necesidad de fármacos tópicos adicionales.
7	En la última exploración oftalmológica, a los 14 meses de vida, se hallaba en ortotropía, con buena fijación y seguimiento de la luz, defecto refractivo de - 4 dioptrías en ambos ojos, ampollas difusas y sin signos inflamatorios, y se confirmaba el buen control de la PIO.
8	En el FO no se encontraron alteraciones de interés.



Task 2 (IR)

Studying consumer health search

Topics

50 topics (same as 2018)

Derived from **query logs** from the Health on the Net (HON) website

Provided with several **spoken version** and their **transcription**

Documents

Webpages from the CommonCrawl for a target domain:

- Top-50 domains retrieved by Bing
- List of trustworthy health pages
- List of not-trustworthy health pages

Assessments

Assessments made by **experts** with respect to:

- Topical Relevance
- Readability
- Credibility

Extension of 2018 pool

Subtask 1: Adhoc IR

Subtask 2: Spoken queries



Wed, Thu, & Fri at 9:00-10:30 AM (Amsterdam time)



Wed 23 Sep: IE & IR

09:00 AM – Welcome & Introduction

09:10 AM – Task 1 Overview

09:25 AM – Task 2 Overview

09:40 AM – Task 1 Participant Presentations (Part 1)

Thu 24 Sep: IE

09:00 AM – Task 1 Introduction & Recap

09:15 AM – Task 1 Participant Presentations (Part 2)

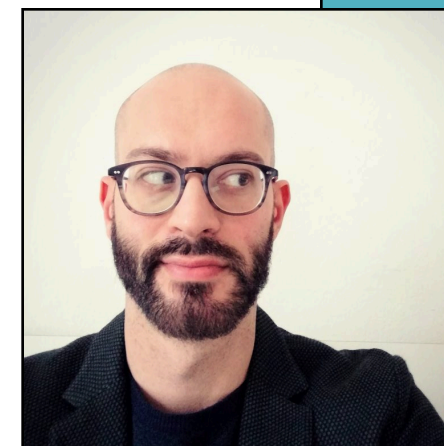
10:20 AM – Task 1 Closing Remarks & Wrapup

Fri 25 Sep: IR

09:00 AM – Keynote: A/Prof Marco Viviani

09:40 AM – Task 2 Participant Presentations

10:25 AM – Task 2 Discussion & Wrap-up



@clefehealth

<https://clefehealth.imag.fr/>



Overview of CLEF eHealth 2012–2020 +2021

Co-chairs: Lorraine Goeuriot, Hanna Suominen, & Liadh Kelly

Task 1: Multilingual IE (Spanish)

Co-leaders: Viviana Cotik & Laura Alonso Alemany

Task 2: Consumer Health Search

Co-leaders: Lorraine Goeuriot, Gabriella Pasi, & Hanna Suominen



*Before we live what's next,
it always seems like there is
some answer we need to arrive at.
But daring to enter,
we are humbled to
**discover, again and again, that
the act of living itself unravels
both the answer and the question.***

Dr Mark Nepo