

Intelligence augmentation for medical diagnosis

Ole Winther

DTU Compute & FindZebra IVS

Joint work with: Dan Svenstrup, Jonas Meinertz Hansen,
Mads Emil Mathiesen and Henrik L Jorgensen



November 20, 2017

Message time!

- Technological progress does not happen by itself!
- Monolithic systems not a good idea
- better give the user freedom to choose between systems.
- Computer systems can be integrated through APIs.

Message time!

- Technological progress does not happen by itself!
- Monolithic systems not a good idea
- better give the user freedom to choose between systems.
- Computer systems can be integrated through APIs.

- FindZebra - a search engine for rare diseases
- Using machine learning to improve search
- Putting text understanding in the hands of users!



Ellen's case story



For 25 years, Ellen struggled to find a diagnosis for the multitude of debilitating symptoms that seemed to increase year after year.

- Her symptoms included muscle cramps, intense headaches, rapid weight gain, fatigue, edema, intolerance to heat, excessive sweating, joint pain, tingling in her hands and feet, frequent bone fractures, acid reflux, intense anxiety and panic attacks, high blood pressure, high cholesterol, high blood sugar, sleep apnea, menstrual irregularities, peripheral vision loss and double vision.
- Source: <http://www.uptodate.com/home/ellen-uses-uptodate-find-diagnosis>
- **Any suggestions?** - Get back to case in demo.

Rare diseases - enter FindZebra.com

“When you hear hoofbeats behind you, don’t expect to see a zebra”



- **Rare diseases** hard to diagnose.
- FindZebra – use latest information technology for faster and more accurate diagnosis
- We want to change the culture of how physicians approach diagnosis

Test queries - examples

- Normally developed boy age 5, progressive development of talking difficulties, seizures, ataxia, adrenal insufficiency and degeneration of visual and auditory functions: ?
- 14 year old, teenage boy, mild mental retardation, proximal muscle weakness, unable to walk (wheelchair-bound), premature ventricular complexes, ophthalmoparesis: ?
- fever, anterior mediastinal mass and central necrosis: ?

Test queries - examples

- Normally developed boy age 5, progressive development of talking difficulties, seizures, ataxia, adrenal insufficiency and degeneration of visual and auditory functions:
Adrenoleukodystrophy autosomal neonatal form
- Ranks: FindZebra=2 and Google search = -
- 14 year old, teenage boy, mild mental retardation, proximal muscle weakness, unable to walk (wheelchair-bound), premature ventricular complexes, ophthalmoparesis:
Autosomal recessive centronuclear myopathy (ARCNM)
- Ranks: FindZebra=2 and Google search = -
- fever, anterior mediastinal mass and central necrosis:
Lymphoma
- Ranks: FindZebra=7 and Google search = 1

Test queries - examples

- Normally developed boy age 5, progressive development of talking difficulties, seizures, ataxia, adrenal insufficiency and degeneration of visual and auditory functions:
Adrenoleukodystrophy autosomal neonatal form
- Ranks: FindZebra=2 and Google search = -
- 14 year old, teenage boy, mild mental retardation, proximal muscle weakness, unable to walk (wheelchair-bound), premature ventricular complexes, ophthalmoparesis:
Autosomal recessive centronuclear myopathy (ARCNM)
- Ranks: FindZebra=2 and Google search = -
- fever, anterior mediastinal mass and central necrosis:
Lymphoma
- Ranks: FindZebra=7 and Google search = 1
- Dragusin et al, 2013: 56 rare disease cases, recall@20:
FindZebra 63% and Google 32%

FindZebra technology - some highlights

FindZebra

Help diagnose rare diseases.

FindZebra is intended primarily for physicians and other professionals concerned with diagnosis of rare diseases. Users seeking information related to personal conditions are urged to consult with a qualified physician. The content of FindZebra is based entirely on publicly available information on rare diseases from highly curated sources. FindZebra cannot be held responsible for any consequences following its use.

- Standard search interface (Google simple to use)
- Extra information with each article (source, genes, drugs)
- Geneview
- Symptom synonyms
- Login functionality
- Dedicated mobile search version

FindZebra search

fever, anterior mediastinal mass and central necrosis

Search

Filters

Group by: disease gene

Mediastinal tumor

Desmoplastic small round cell ... (2)

Thymoma (2)

Adult-onset Still's disease (3)

Large-cell lymphoma (3)

Follicular lymphoma

Periodic fever, familial, autoso... (2)

Thymic carcinoma

TNF receptor associated periodic...

Japanese encephalitis

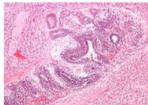
Periodic fever, familial, autosomal...

Mediastinal tumor

Retrieved: 28-09-2014

Source: WIKIPEDIA ([Original article](#))

The mediastinum is the cavity that separates the lungs from the rest of the chest. It contains the heart, esophagus, trachea, thymus, and aorta. The mediastinum has three main parts: the **anterior** mediastinum (front), the middle mediastinum, and the posterior mediastinum (back). The most common **mediastinal** masses are neurogenic tumors (20% of **mediastinal** tumors), usually found in the posterior mediastinum, followed by thymoma (15-20%) located in the **anterior** mediastinum. Masses in the **anterior** portion of the mediastinum can include thymoma, lymphoma, pheochromocytoma, germ cell tumors including teratoma, thyroid tissue, and parathyroid lesions. Masses in this area are more likely to be malignant than those in other compartments. Masses in the posterior portion of the mediastinum tend to be neurogenic in origin, and in adults tend to be of neural sheath origin including neurilemmomas and neurofibromas. Lung cancer typically spreads to the lymph nodes in the mediastinum.



Diagnosis

In several editions of Physical Diagnosis, concerning **mediastinal** tumors the author writes: According to Christian1 the **mediastinal** neoplasms which are neither so rare nor so obscure as to make diagnosis practically impossible are:

FindZebra extra information

fever, anterior mediastinal mass and central necrosis

Search

Filters

Group by: disease gene

Mediastinal tumor

Desmoplastic small round cell ... (2)

Thymoma (2)

Adult-onset Still's disease (3)

Large-cell lymphoma (3)

Follicular lymphoma

Periodic fever, familial, autos... (2)

Thymic carcinoma

TNF receptor associated periodic...

Japanese encephalitis

Periodic fever, familial, autosomal...

Follicular lymphoma

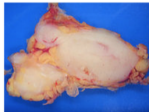
Retrieved: 28-09-2014

Source: Orphanet ([Original article](#))

Associated genes: [h1a-drb1](#), [igh](#), [bcl2](#), [bcl6](#)

Market authorized treatments: (S)-3-(1-(9H-purin-6-ylamino)ethyl)-8-chloro-2-phenylisoquinolin-1(2H)-one, 177Lu-tetraxetan-tetolomab, 2-[(1R,6R)-3-Methyl-6-(1-methylethyl)-2-cyclohexen-1-yl]-5-pentyl-1,3-benzenediol, 5-aminolevulinic acid hydrochloride, Bendamustine for 50 ml admixture, Doxorubicin hydrochloride (pegylated liposomal), Gabapentin enacarbil, Human coagulation factor VIII / Von Willebrand factor, Ibritumomab tiuxetan, Interferon alpha 2b, Lucinactant, Mepolizumab, Rifampicin, Isoniazid / pyrazinamide, Rituximab, S[+] apomorphine, Temozolomide, Tositumomab, Trans-4-[4-[5-[6-(trifluoromethyl)-3-pyridinyl]amino]-2-pyridinyl]phenyl] cyclohexane acetic acid sodium salt

Follicular lymphoma is a form of non-Hodgkin lymphoma (see this term) characterized by a proliferation of B cells whose nodular structure of follicular architecture is preserved. Prevalence of follicular lymphoma is estimated at about 1/3,000. The median age at diagnosis is 60-65 years. The disease is extremely rare in children. Follicular lymphoma is located primarily in the lymph nodes, but can also involve the spleen, bone marrow, peripheral blood and Waldayer's ring. The skin and **central** nervous system are affected in rare cases. Symptoms appear at an advanced stage of the disease and can include **fever**, night sweats and weight loss. At diagnosis, patients usually present with adenopathy and, in 50% of cases, splenomegaly. In 85% of cases, follicular lymphomas



FindZebra geneview

fever, anterior mediastinal mass and central necrosis

Search

Filters

Group by: disease gene

[Download gene list...](#)

EWSR1

WT1

BCL2 (B-cell CLL/lymphoma 2)

IGH (immunoglobulin heavy locus)

BCL6 (B-cell CLL/lymphoma 6)

HLA-DRB1

TNFR1

FPF

TNFAR

MS5

TNFRSF1A

EWSR1

Symbol: EWSR1

Synonyms: None

External references: [omim](#)

Associated diseases (among search results):

- [Desmoplastic small round cell tumor](#)
- [Desmoplastic small round cell tumor](#)

FindZebra synonyms

Filters

Group by: disease gene

Periodontosis

Periodontitis, chronic

Periodontitis, aggressive, 1

Ehlers-danlos syndrome, type viii

Periodontosis

Retrieved: 28-09-2014

Source: WIKIPEDIA ([Original article](#))

Periodontosis is an obsolete term that was used to describe what was once thought to be certain type of unique and distinguishable chronic periodontal disease that manifested as degenerative bony changes without concomitant inflammation. Although utilized for more than 50 years, the term has since been dropped in favor of a more contemporary disease classification for periodontal disease. Described by Gottlieb as a "diffuse atrophy of the alveolar bone," the term **periodontosis** was later applied and it gained acceptance as a disease entity, Periodontal disease in children. In Goldman HM, Cohen DW, editors: Periodontal Therapy, 5th Edition. St. Louis: Mosby, Inc. 1973. page 259-274 being defined as: "a degenerative, noninflammatory destruction of the periodontium, originating in one or more of the periodontal structures and characterized by migrating and loosening of the teeth in the presence or absence of secondary epithelial proliferation and pocket formation or secondary gingival disease." Noted as a rare disease, **periodontosis** was said to have been seen primarily in young patients. And despite being defined as being a "noninflammatory destruction of the periodontium," almost all cases did exhibit varying degrees of gingival inflammation.

Improving search w good data and machine learning

- FindZebra uses high quality sources (NIH, OMIM, OrphaNet, Wikipedia)
- 14k+ articles on 8k+ diseases (UMLS)

Improving search w good data and machine learning

- FindZebra uses high quality sources (NIH, OMIM, OrphaNet, Wikipedia)
- 14k+ articles on 8k+ diseases (UMLS)
- **Expand dataset** by
 - Search w disease name and extract data from found domains (eg. WebMD)
 - Use PubMed Central Open Access case studies with disease name in title.

Improving search w good data and machine learning

- FindZebra uses high quality sources (NIH, OMIM, OrphaNet, Wikipedia)
- 14k+ articles on 8k+ diseases (UMLS)
- **Expand dataset** by
 - Search w disease name and extract data from found domains (eg. WebMD)
 - Use PubMed Central Open Access case studies with disease name in title.
- **Machine learning** is data hungry:
 - Use deep learning models
 - Train on text snippets.

Improving search w good data and machine learning

- FindZebra uses high quality sources (NIH, OMIM, OrphaNet, Wikipedia)
- 14k+ articles on 8k+ diseases (UMLS)
- **Expand dataset** by
 - Search w disease name and extract data from found domains (eg. WebMD)
 - Use PubMed Central Open Access case studies with disease name in title.
- **Machine learning** is data hungry:
 - Use deep learning models
 - Train on text snippets.
- Recall@20 performance on 250 test queries:

	Baseline	Extended	Pubmed
Neural BOW (without synonyms)	62.90	70.97	77.42
Neural BOW (with synonyms)	63.31	71.37	77.42
Neural n-gram	48.79	60.48	64.52
Word LSTM	46.97	58.06	66.94
Char LSTM	46.37	64.11	63.71
Solr (with synonyms)	68.55	69.76	76.41
Solr (without synonyms)	68.55	68.75	75.81

- Ensemble of best two: **81.05%**.

How can we get digital innovation into healthcare?

- We can improve diagnosis with machine learning driven hypothesis generation and
 - add AI to patient-healthcare system interaction, however
- ① How to fit into the current very set system?
 - ② Who will pay for digital healthcare solutions?
 - ③ What is the best set-up for pilot studies and distribution?

How can we get digital innovation into healthcare?

- We can improve diagnosis with machine learning driven hypothesis generation and
 - add AI to patient-healthcare system interaction, however
- ① How to fit into the current very set system?
 - ② Who will pay for digital healthcare solutions?
 - ③ What is the best set-up for pilot studies and distribution?
- `findzebra.com` - a search engine for rare diseases
 - Large potential for tools in the patient-healthcare system interface
 - Augment human intelligence - do not replace it.



Thank you!

olwi@dtu.dk