

mining unstructured healthcare data

deep dhillon

chief data scientist | ddhillon@alliancehealth.com | twitter.com/zang0

alliance health networks

HCP/advanced patients
 medify.com – 10,000% user growth past year

patients
 diabeticconnect.com - # 1 online diabetes site @ ~1.4M uniques
 *connect.com – content sites + guided social networks

health care industry
 patient surveying, matchmaking and analysis

why mine healthcare text?

anatomy of what patients currently use: webmd (e.g. drugs.com, yahoo health, etc.)

q/a	topic pages	news + media	experts	discussions
<ul style="list-style-type: none"> original content aligns well with google searches provides answers 	<ul style="list-style-type: none"> original content typically aligned well with google searches, i.e. treatments for X, symptoms for Y good coverage in the head 	<ul style="list-style-type: none"> original content fresh simple to understand good coverage in the head 	<ul style="list-style-type: none"> original content moderately authoritative simple to understand good coverage in the head provides answers 	<ul style="list-style-type: none"> original content addresses emotional needs simple to understand provides answers

original content=important
 providing answers=important
 head=developed
 fresh=important
 authority=important

why mine healthcare text?

Patients and HCPs need long tail, statistically meaningful, consumer friendly, authoritative and fresh health content.

q/a	topic pages	news + media	experts	discussions
<ul style="list-style-type: none"> manually written, not authoritative not consistently credible, i.e. minimal accreditation not statistically meaningful sparse 	<ul style="list-style-type: none"> evergreen, rarely change manually written, not authoritative not consistently credible, i.e. minimal accreditation not thorough, i.e. not long tail 	<ul style="list-style-type: none"> manually written, not authoritative not consistently credible, i.e. minimal accreditation not thorough, i.e. not long tail 	<ul style="list-style-type: none"> manually written, moderately authoritative not thorough, i.e. not long tail 	<ul style="list-style-type: none"> manually written, not authoritative not thorough, i.e. not long tail

manual=expensive
 manual=head focused
 manual=not authoritative
 manual=old and dated

automated content generation

- cost effective
- structured content
- statistically based
- scales to millions of patients
- scales to long tail treatments + conditions
- authoritative / citation driven
- fresh

types of text mining

Document Classification

USPS - Fast Delivery Shipping 1-4 day USA
 Best quality drugs
 Worldwide shipping
 Professional packaging
 100% guarantee on delivery
 Best prices in the market
 Discounts for returning customers
 FDA approved products
 350000+ satisfied customers

Model

Spam

Not Spam

Information Extraction

We retrospectively explored changes in immunological parameters in men with biochemically recurrent prostate cancer treated with either 5 or 25 mg of lenalidomide in a randomized phase 2 trial, and determined whether those changes correlated with disease progression.

Extractor

recurrent prostate cancer

5 or 25 mg

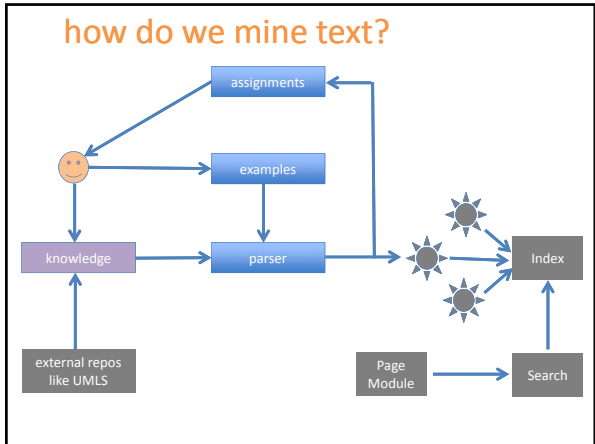
lenalidomide

NLP Grammatical Parsing

Fruit flies like a banana

Parser

medify demo

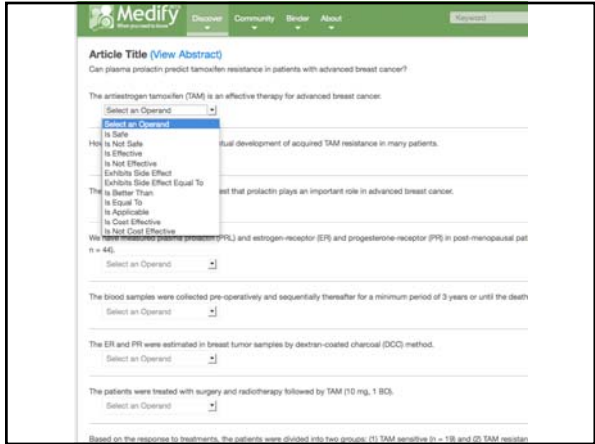
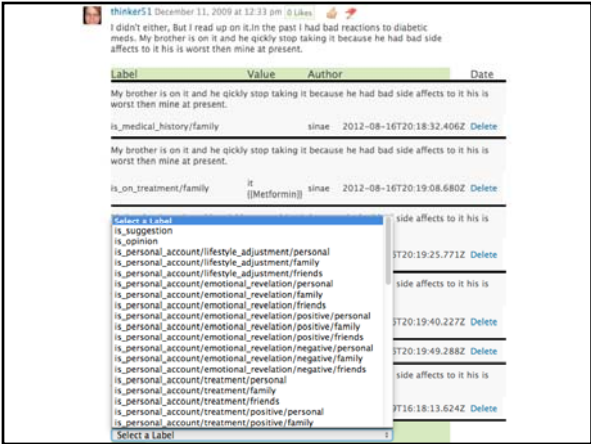


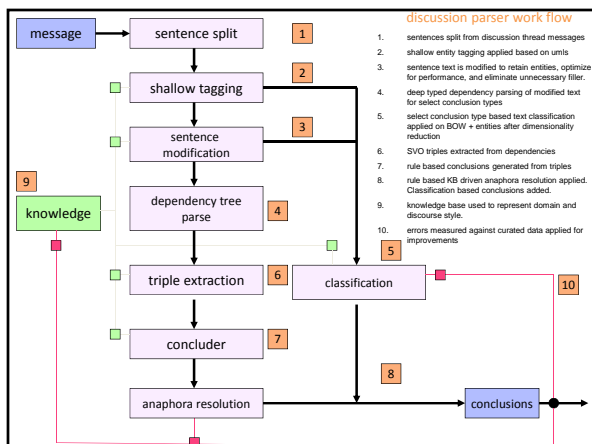
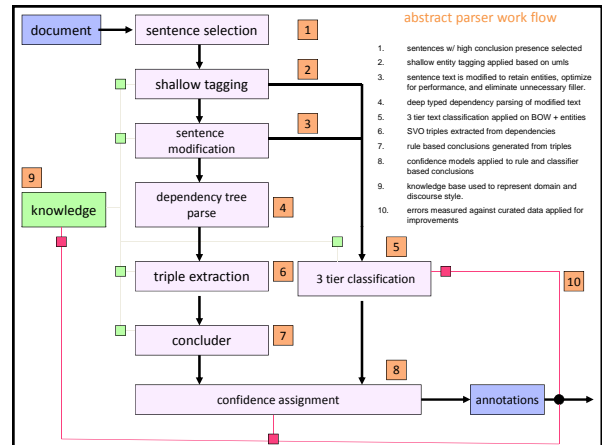
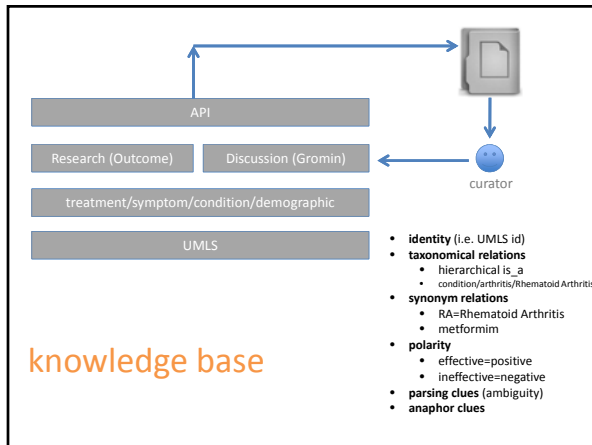
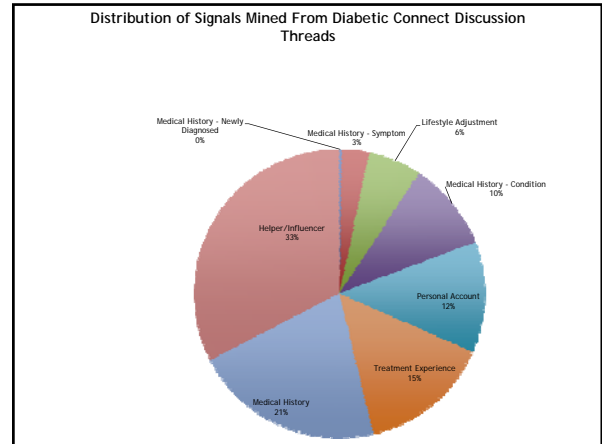
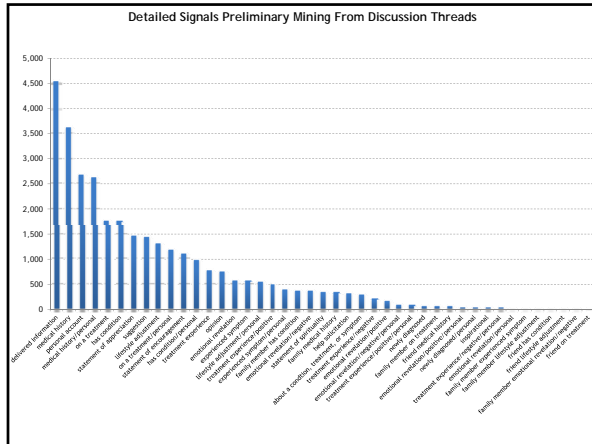
annotate in the product:

- Easier for people to give explicit GT feedback
- Channels high visibility error angst productively
- Channels more GT toward areas most seen by users
- Override high visibility system mistakes with human data

gt demo:

- <http://www.diabeticconnect.com/discussions/4790>
- <https://www.medify.com/internal/annotate/abstract?abstractid=8181575>





- ### feature engineering
- entities, i.e. normalize synonyms, id new entity types, like social relations
 - entity types, i.e. metformin > treatment_medication
 - phrase driven cues, i.e. [have] [you] [considered] > suggestion_indicator

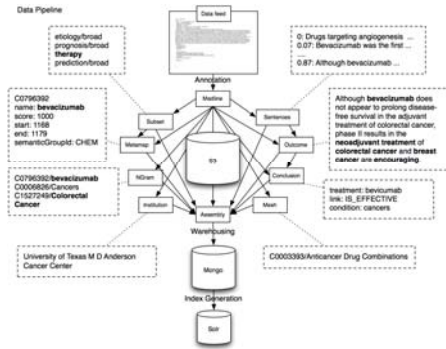
anaphora resolution

- relation structure, i.e. [person]>takes>it
it refers to treatment (i.e. not condition/symptom), and specifically: medication but not device
- statistically driven, manually curated cues
i.e. drug > treatment/medication
- filter
 - non matching antecedent candidates
 - singular/plural agreement
- score candidates:
 - antecedent occurrence frequency
 - distance (#sents) from antecedent to anaphor
 - co-occurrence of anaphor w/ antecedent

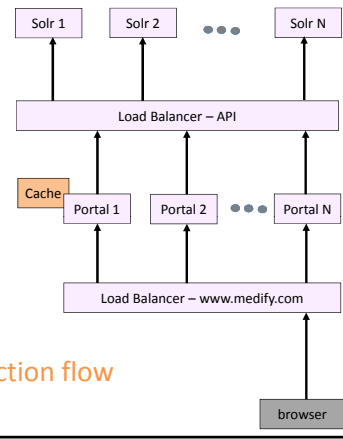
technology

- Lang: Java + Python + Ruby
- DB: Solr 4, Mongo DB, S3
- Work: Map Reduce
- Dependencies: Malt Parser, Stanford Parser
- Misc: Tomcat, Spring, Mallet, Reverb, Minor Third
- Tagging: Peregrine + home grown

Data Pipeline



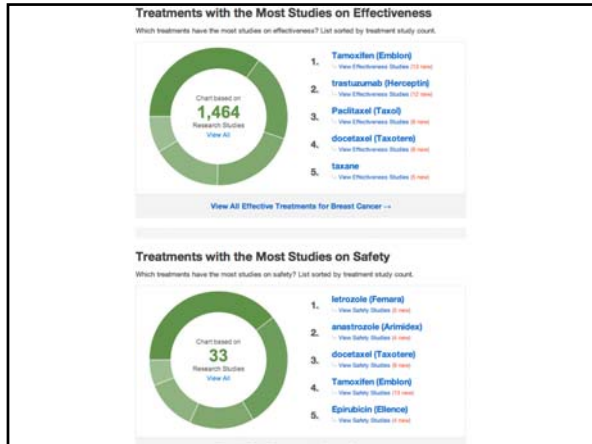
21



request transaction flow

questions?

The screenshot shows a web interface for breast cancer information. It features a 'Breast Cancer' header with a definition and statistics. Below this, there are sections for 'What Health Experts Say', 'What Patients Say', and 'Trending Treatments'. The trending treatments list includes Adjuvant Chemotherapy (10,371 studies), Radiation Therapy (4,282 studies), Neoadjuvant Therapy (2,112 studies), trastuzumab (Herceptin) (1,883 studies), and Tamoxifen (Emblon) (9,255 studies). On the right, there are sections for 'Questions', 'Medify Insights', and 'Community'.



Display studies where Tamoxifen is effective for Breast Cancer

Breast Cancer

Date	Patients	Outcome
1994-06-14	150	The study clearly demonstrated that plasma protein levels indicated the response and development of resistance to TAM in patients with advanced breast cancer.
1999-04-23	150	Tamoxifen has some weak estrogenic activity.
2003-06-26	150	Tamoxifen improved survival and DFS.
1999-03-19	150	Tamoxifen alone may be of value.
1991-04-26	150	Tamoxifen prevents relapse of endometrial carcinomas.
2004-07-26	150	Tamoxifen also demonstrates some weak estrogenic activity.
1993-09-16	150	Tamoxifen is the systemic treatment of choice for breast cancer.

Abstract Summary

Prognostic Factor: The general personal and health characteristics that have a bearing on the course of a disease and its eventual outcome. [SOURCE: UMLS]

Full Abstract

Full Article

Breast Cancer

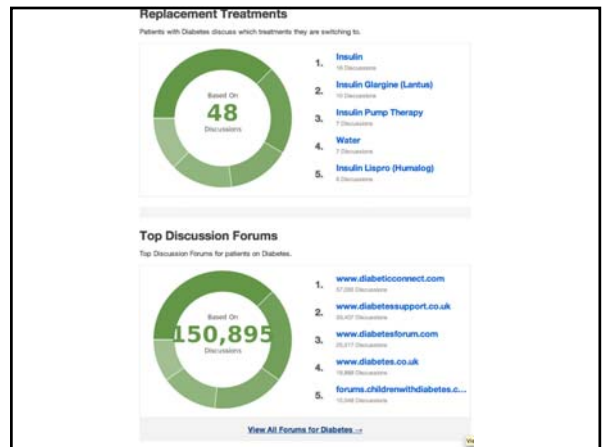
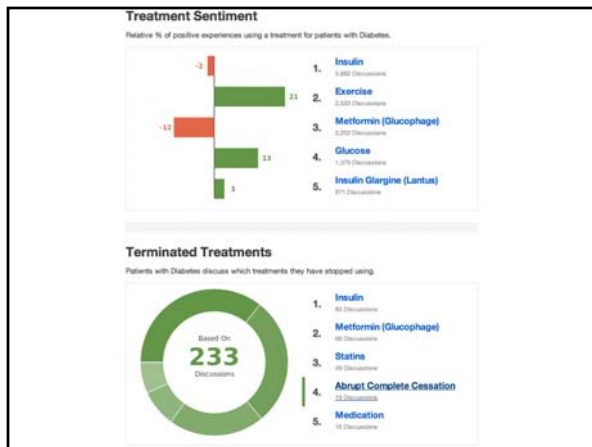
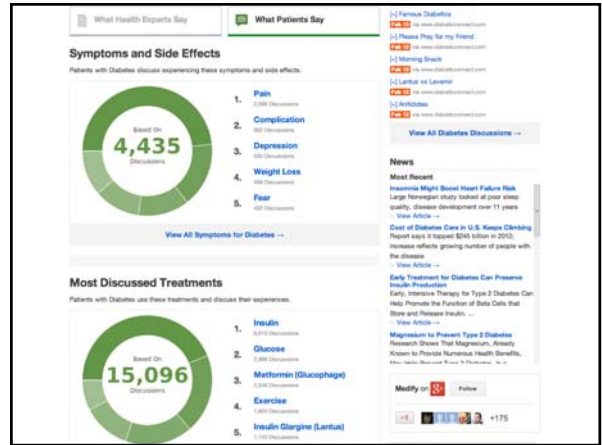
... of this study was to identify the patients who are at low or high-risk by defining the prognostic factors in node-negative breast carcinomas.

METHODS: Medical records of 384 consecutive breast cancer patients with negative axillary lymph nodes who had been operated on between January 1994 and January 1997 at our hospital were retrospectively reviewed. Surviving clinical and pathological characteristics of patients were categorized. Univariate analyses of survival and disease-free survival (DFS) were performed by the Kaplan-Meier method and the log-rank test. Independent prognostic and predictive factors affecting survival and DFS were assessed by Cox proportional hazard method.

RESULTS: 5-year survival and DFS were 91.4 and 86.7%, respectively. Sex, grade, age, and lymphovascular invasion (LVI) were the prognostic factors that independently affected survival and DFS. Tamoxifen improved survival and DFS. While age younger than 35 was an adverse factor for both survival and DFS, age older than 49 was a detrimental factor for DFS.

CONCLUSIONS: Patients who have a tumor with size greater than 2 cm, with histologic grade 3, with LVI, and patients with age under 35 or older than 49 have poorer prognosis among node-negative breast carcinomas, and are candidates for adjuvant therapy.

From MEDLINE®/PubMed®, a database of the U.S. National Library of Medicine. See article source on PubMed.



Get discussion | messages | with patients | speaking negatively about | Metformin

Displaying 77 messages with patients speaking negatively about Metformin

Add Filters
Applied Filters: Diabetes | Custom Keyword

Date	Outcome
-	What I WILL tell you is that after trading my diabetes for three years & going a few rounds with me over what works and what doesn't, my own doctor, a highly regarded physician of Merry Chase of southeast Missouri (president of the Falls Division of Merit), a graduate of Vanderbilt School of Medicine, and board certified specialist in the Internal Medicine sub-specialty of Geriatrics, is persuaded that natural saturated fats and monounsaturated fats constitute a healthy diet. ... You then proceeded to tell me I needed to check my facts because metformin can cause lactic acidosis. [more]
-	I am "benign" diabetic. A1c was 7.4. ... I don't know if metformin causes depression, but it wouldn't amaze me to hear that it does. [more]
-	Metformin may commonly cause stomach upset, indigestion, nausea, vomiting, or diarrhea at the beginning of treatment. [more]
-	I was honeymoon type 1 as a child, and am now type 1, but, because of internal bleeding, and not eating sugar for 7 whole years, my A1C actually dropped way LOW, after METFORMIN ALMOST KILLED ME. ... [more]

Get discussion | messages | with patients | speaking positively about speaking negatively about quitting switching to suffering from condition suffering from symptom changing diet & exercise medical history about Metformin

Displaying 77 messages with patients speaking

Add Filters
Applied Filters: Diabetes | Custom Keyword

Metformin (Glucophage)

Definition: Metformin is an oral antidiabetic drug in the biguanide class. It is the first-line drug of choice ... more

30,574 Discussions (View)

What Health Experts Say | **What Patients Say**

Symptoms and Side Effects

Patients using Metformin (Glucophage) discuss experiencing these symptoms and side effects.

Based On **309** Discussions

- Pain** - 302 Discussions
- Weight Loss** - 184 Discussions
- Headache** - 116 Discussions
- Nausea** - 80 Discussions
- Muscle Cramp** - 77 Discussions

[View All Symptoms and Side Effects for Metformin \(Glucophage\)](#)

Get discussion | messages | with patients | suffering from symptom | Headache

Displaying 116 messages with patients suffering from symptom Headache

Add Filters
Applied Filters: Metformin | Custom Keyword

Date	Outcome
-	I have not had any need for ibuprofen aspirine (almost two to four [500 mg] daily) since my headaches has stopped. ... When my doctor passed me off by her facial expression showing doubt when I vowed to get off the insulin and hopefully Metformin, I was more determined than ever to make the change. [more]
-	I get really bad headaches or migraines sometimes and the only thing that will get rid of them is coffee. ... I am only on metformin so maybe that's why it affects us different? ... [more]
-	I'm now taking metformin as I want to lose weight and get to a normal blood sugar a bit faster than diet alone. ... Since I reduced my readings I no longer get regular headaches or fatigue - I have a lot more energy. [more]

Quote: Originally Posted by MMS1 How long do you feel you have to be on Xanax and at what dose before you seriously run into this problem? I've been taking a .5 mg dose for about seven weeks due to high anxiety and depression (due to a hearing loss in Feb.) and not being able to sleep since then. [usa]

[View Thread](#)
[Discussed on www.healthboards.com by pursia](#)
[Topics: Panic Disorder, Alprazolam, Depression](#)

Advair: Experts vs. Patients

Condition	Expertise Rate	Patient Rate
1. Common Cold	18 - 27%	100%
2. Pharyngitis	13%	100%
3. Candidiasis	3 - 10%	100%
4. Respiratory Infe...	3 - 6%	100%
5. Nausea	4 - 5%	100%

- "medicalese" vs. patients words
- more granularity
- a story like perspective w/ words of inspiration

Based On **177** Discussions

- Pain** - 60 Discussions
- Cough** - 48 Discussions
- Flare** - 32 Discussions
- Hurt** - 20 Discussions
- Chest Pain** - 15 Discussions

My Pulmonologist today said that he had just come from the hospital bedside of a patient with my exact symptoms who is on oxygen and not doing well. If I hadn't been so diligent in following my asthma plan that could easily have been me. His telling me that really hit home.

By the way, my Pulmonologist surprised me with his view on many of his patients. He actually got excited and thanked me for being knowledgeable about asthma, understanding my own body and health needs, and for following my treatment plan. He said he doesn't get many patients who can actually talk about their disease, symptoms, time frames, treatments, medications, and non-medical measures they are taking. He also doesn't get many people who ask questions. My response to this: What are people thinking? They need to take control of their own health or noone else will.