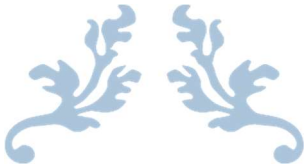

DIABETES ENGINE

ARCHITECTURE

MARCH 13, 2017

MEDWHAT
Akshay Surendra Phadnis

Table of Contents

Diabetes Engine Schematic

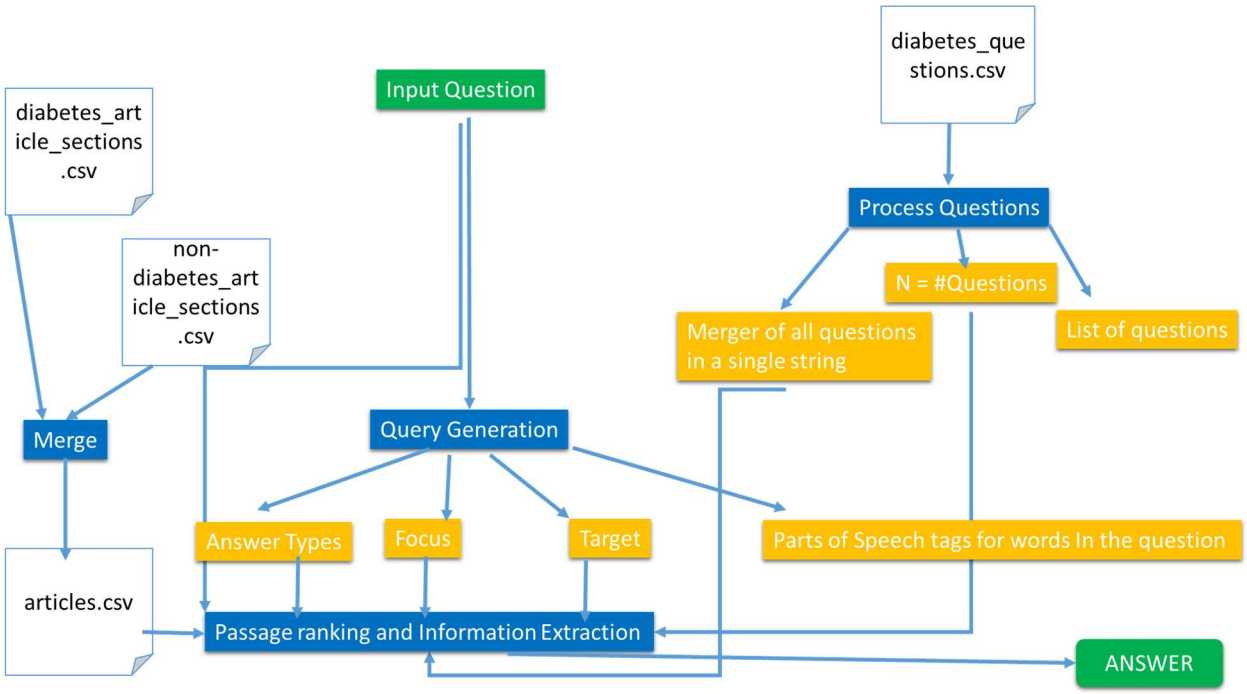

Fig. 1: Diabetes Engine Schematic

Processing diabetes questions

A file replete with common diabetes questions is read and each question is stored, individually and as a part of a string that will hold all the questions as single string. The number of questions encountered will also be stored.

How does this help?

The diabetes engine makes use of synonyms of various words from questions while crafting the query. However, not all words can be used. So, we keep only those synonyms that pertain to medical parlance, especially regarding diabetes. This is done be computing IDF (inverse document frequency) scores for each candidate synonym and keep only those that have an IDF score greater than 1. This is explained in detail in the sub section on answer types.

Query generation:

In-order to probe the CSV data file (formed by combining data -CSV files-about diabetes and about diseases other than diabetes), a query must be formed that tell us what to search for.

This problem is viewed as follows:

Every question is a string and it has a primary topic called focus. The question seeks information about existence of a relation f such that

f(focus,target) is true , where target is a group of words (could be considered as a single string) that has the relationship f (which will be the answer type) with the focus.

For instance,

Cause(diabetes, x) means we are looking for a candidate passage about causes of diabetes. Stated other way, we check the causes sub section of a passage about diabetes. If it contains x then the passage is also about the same relation, focus and target as the question is and hence might be a probable answer.

As another example:
i eat crazy amounts of sweets. so is it possible to get diabetes from this
ANS TYPE: causes
FOCUS: diabetes
TARGET: i crazy amounts sweets possible get
here, we look for passages about causes of diabetes such that they mention sweets or amounts of sweets, thereby including this semantic "do sweets cause diabetes"

The query generated will be a tuple containing

(ans types, focus, target, parts of speech tags for each word in the question)

Focus and Target:

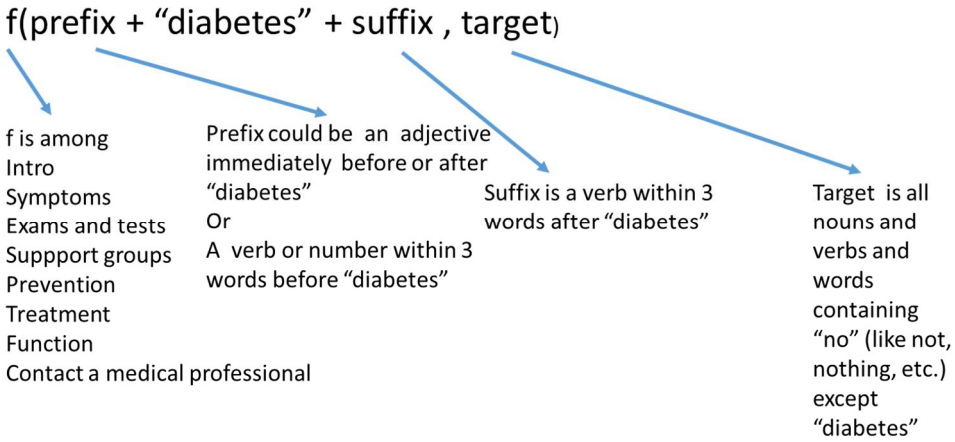

Fig. 2: Focus and Target

Ans type detection:

Answer type is determined using regular expressions. Every time a word among those listed for f above or their synonyms is detected, that word becomes the answer type and for a given question, there can be multiple of those. This redundancy is deliberate. It allows ranking more passages and thus increases the chances of finding better answers.

For instances, the answer type “cause” is detected as follows:

```
syn_causes = obtain_synonyms("causes")
   re_cause = ".+(?:cause)"
   for w in syn_causes:
      if idf(w,N,common_diabetes_questions) > 1:
         re_cause =  re_cause + "|(?:"+ st.stem( w.replace("_", " ") ) + ")"
   re_cause = re_cause + ".+"
if findall(re_cause,q ) != []:
      ans_types.append( "causes" )
```

The resulting regular expression would look like:
u'.+(?:cause)|(?:caus)|(?:caus)|(?:reason)|(?:ground)|(?:caus)|(?:driv)|(?:mov)|(?:caus)|(?:lawsuit)|(?:suit)|(?:cas)|(?:caus)|(?:caus)|(?:caus)|(?:mak)|(?:induc)|(?:caus)|(?:get)|(?:mak).+
The regular expression stores stems or base forms for words instead of complete words. This is done so that variations like ‘causes’ and ‘causing’ are recognized as being about ‘causes’.

Since the number of synonyms being considered could be too large and not all of them could pertain to diabetes, we keep only those that do. How to assess this? One way would be to keep those synonyms that have the highest frequency in a corpus pertaining to diabetes. But this approach has a flaw. For instance, the word “cause” has a synonym “do” which can appear with high frequency, even in a corpus that is about diabetes (and not English literature).

To avoid this issue instead of choosing synonyms with high frequency, we choose those with high IDF (inverse document frequency). IDF for a word w is defined as:
log(N/d)
N = Number of documents
d = number of documents in which w appears.

For this implementation, I have used commonly asked diabetes questions as my corpus such that the document in the above definition corresponds to a question. Thus,
N = Number of questions
d = number of questions in which w appears

A high IDF score means N/d is high (compared to IDF score of some other word). This means that those words which d is less i.e. w is a rare occurrence which would be true for medical terms and words like “do” which appear frequently would have a lower IDF score and would hence get rejected.

Information Retrieval

By this stage the program is armed with focus, target and ans types. The input data file, a labelled CSV in this case is scanned line by line.

For each row,

If it's article title or any synonyms of words in the title have even a partial match with the focus of the question, the article in that row is considered as a candidate passage.

This however isn't narrow enough for searching the answer. The ans type from the question further assists us in doing that. If there is even a partial match between ans type of the question and the section title of the row, we have more confidence that this article is indeed a candidate passage. Otherwise, that passage is not considered as a candidate passage.

What this does is, If I have the question "how to prevent gestational diabetes", among several rows regarding gestational diabetes, only the row that talks about prevention is considered as a candidate passage.

It would seem, that performing only these two techniques would land us to the perfect answer, however that does not happen. This is because there is always a chance that the focus and target detected for the question do not have an exact match to the article title or the section title.

As a result we get several candidates following the same relation f(focus, target).

Among these candidates, the best needs to be chosen. This is achieved by scoring each passage and choosing the one with the highest score.

This score is comprised is calculated as:

10*num_common_words + CTA+ P + overlap_focii + overlap_targets+ overlap_anstypes

Where,

- num_common_words is number of common words between title words (including synonyms of each word) and focus words (including synonyms and stems of each word)
- CTA is the number of common words between article and target words (including synonyms and stems of each word)
- P denotes how close are the words from the question inside a given passage and is called a proximity score

Focus detection and target recognition are also applied to each passage. The idea here is that the answer to question would entail similar focus, target and ans types as the question and thus, computing the number of common words or overlap between the two foci or the two targets or the two ans types

are a good metric for scoring a candidate passage. This is captured by the remaining features or metrics used for computing the score.

Resolving ties

There might be more than one passage with the highest score, in these cases, the one where proximity is lower (tightness -in terms of closeness of words from question- is high) is chosen.

Effect of type detection on candidate passages

With ans type detection even on candidate passages to check if the question and the candidate passage refer to the same thing, the accuracy has improved, the answers are crisper, more specific: for instance:

1.

Previously :

i'm a fit 45 year old male who sometimes smokes. my immediate family history includes diabetes and heart disease. what tests should i ask my doc for

ANS TYPE: tests

FOCUS: diabetes

TARGET: i fit year old male smokes immediate family history heart disease i ask doc

CHOSEN ANS:

Your health care provider may suspect that you have diabetes if your blood sugar level is higher than 200 mg/dL. To confirm the diagnosis, one or more of the following tests must be done. Diabetes blood tests: Fasting blood glucose level -- diabetes is diagnosed if it is higher than 126 mg/dL two times
Hemoglobin A1C test --Diabetes: 6.5% or higher Oral glucose tolerance test -- diabetes is diagnosed if glucose level is higher than 200 mg/dL 2 hours after drinking a special sugar drink Diabetes screening is recommended for: Overweight children who have other risk factors for diabetes, starting at age 10 and repeated every 2 years
Overweight adults (BMI greater than 25) who have other risk factors
Adults over age 45 every 3 years If you have been diagnosed with type 2 diabetes, youneed to work closely with your health care provider. You will likely need to see your provider every 3 months. At these visits, you can expect your provider to do the following: Check your blood pressure
Check the skin and bones on your feet and legs
Check if your feet are becoming numb
Examine the back part of the eye with a special lighted instrument The following tests will help you and your doctor monitor your diabetes and prevent problems: Have your blood pressure checked at least every year (blood pressure goals should be 140/80 mm/Hg or lower).
Have your A1C test (hemoglobin A1C) every 6 months if your diabetes is well controlled; otherwise every 3 months.
Have your cholesterol and triglyceride levels checked yearly (aim for LDL levels below 70-100 mg/dL).
Get yearly tests to make sure your kidneys are working well (microalbuminuria and serum creatinine).
Visit your eye doctor at least once a year, or more often if you have signs

of diabetic eye disease.
See the dentist every 6 months for a thorough dental cleaning and exam. Make sure your dentist and hygienist know that you have diabetes.

Now: i'm a fit 45 year old male who sometimes smokes. my immediate family history includes diabetes and heart disease. what tests should i ask my doc for

ANS TYPE: tests

FOCUS: diabetes

TARGET: i fit year old male smokes immediate family history heart disease i ask doc

CHOSEN ANS:

diabetes is diagnosed with the following blood tests: fasting blood glucose level -- diabetes is diagnosed if it is higher than 126 mg/dl two times
random (nonfasting) blood glucose level -- you may have diabetes if it is higher than 200 mg/dl, and you have symptoms such as increased thirst, urination, and fatigue (this must be confirmed with a fasting test)
oral glucose tolerance test -- diabetes is diagnosed if the glucose level is higher than 200 mg/dl 2 hours after drinking a special sugar drink
hemoglobin a1c test -- 6.5% or higher ketone testing is also sometimes used. the ketone test is done using a urine sample or blood sample. ketone testing may be done: when the blood sugar is higher than 240 mg/dl
during an illness such as pneumonia, heart attack, or stroke
when nausea or vomiting occur
during pregnancy the following tests or exams will help you and your doctor monitor your diabetes and prevent problems caused by diabetes: check the skin and bones on your feet and legs.
check to see if your feet are getting numb.
have your blood pressure checked at least every year (blood pressure goal should be 140/80 mm/hg or lower).
have youra1c test (hemoglobin a1c) done every 6 months if your diabetes is well controlled; otherwise, every 3 months.
have your cholesterol and triglyceride levels checked yearly (aim for ldl cholesterol levels below 70-100 mg/dl).
get yearly tests to make sure your kidneys are working well (microalbuminuria and serum creatinine).
visit your eye doctor at least once a year, or more often if you have signs of diabetic eye disease.
see the dentist every 6 months for a thorough dental cleaning and exam. make sure your dentist and hygienist know that you have diabetes.

2.

Previously: what test do doctors conduct to see if i have gestational diabetes

ANS TYPE: treatment tests

FOCUS: gestational diabetes

TARGET: test doctors conduct see i gestational

CHOSEN ANS:

The goals of treatment are to keep blood sugar (glucose) levels within normal limits during the pregnancy, and to make sure that the growing baby is healthy. WATCHING YOUR BABY Your health care provider should closely check both you and your baby throughout the pregnancy. Fetal monitoring will check the size and health of the fetus. A nonstress test is a very simple, painless test for you and your baby. A machine that hears and displays your baby's heartbeat (electronic fetal monitor) is placed on your abdomen. Your health care provider can compare the pattern of your baby's heartbeat to movements and find out whether the baby is doing well. DIET AND EXERCISE The best way to improve your diet is by eating a variety of healthy foods. You should learn how to read food labels, and check them when making food decisions. Talk to your doctor or dietitian if you are a vegetarian or on some other special diet. In general, when you have gestational diabetes your diet should: Be moderate in fat and protein
Provide your carbohydrates through foods that include fruits, vegetables, and complex carbohydrates (such as bread, cereal, pasta, and rice) Be low in foods that contain a lot of sugar, such as soft drinks, fruit juices, and pastries If managing your diet does not control blood sugar (glucose) levels, you may be prescribed diabetes medicine by mouth or insulin therapy. Most women who develop gestational diabetes will not need diabetes medicines or insulin, but some will.

Now: what test do doctors conduct to see if i have gestational diabetes

ANS TYPE: treatment tests

FOCUS: gestational diabetes

TARGET: test doctors conduct see i gestational

CHOSEN ANS:

gestational diabetes usually starts halfway through the pregnancy. all pregnant women should receive an oral glucose tolerance test between the 24th and 28th week of pregnancy to screen for the condition. women who have risk factors for gestational diabetes may have this test earlier in the pregnancy. once you are diagnosed with gestational diabetes, you can see how well you are doing by testing your glucose level at home. the most common way involves pricking your finger and putting a drop of your blood on a machine that will give you a glucose reading.

Here, this answer, one can notice the improvement by noticing that the second answer clearly mentions glucose tolerance tests as an exam for gestational diabetes.

Because of the type detection on candidate passages, answers to even incorrect answers seem are those that are about the focus and target of the question for instance:

Previously: do you experience double vision when you have type 2 diabetes

ANS TYPE: intro

FOCUS: type 2 diabetes

TARGET: do experience double vision

CHOSEN ANS:

Until recently, the common type of diabetes in children and teens was type 1. It was called juvenile diabetes. With Type 1 diabetes, the pancreas does not make insulin. Insulin is a hormone that helps glucose,or sugar, get into your cells to give them energy. Without insulin, too much sugar stays in the blood. But now younger people are also getting type 2 diabetes. Type 2 diabetes used to be called adult-onset diabetes. But now it is becoming more common in children and teens, due to more obesity. With Type 2 diabetes, the body does not make or use insulin well. Children have a higher risk of type 2 diabetes if they are obese, have a family history of diabetes, or are not active, and do not eat well. To lower the risk of type 2 diabetes in children Have them maintain a healthy weight
Be sure they are physically active
Have them eat smaller portions of healthy foods
Limit time with the TV, computer, and video Children and teens with type 1 diabetes may need to take insulin. Type 2 diabetes may be controlled with diet and exercise. If not, patients will need to take oral diabetes medicines or insulin.

Now: do you experience double vision when you have type 2 diabetes

ANS TYPE: intro

FOCUS: type 2 diabetes

TARGET: do experience double vision

CHOSEN ANS:

type 2 diabetes is a lifelong (chronic) disease in which thereis ahigh level of sugar (glucose) in the blood. type 2 diabetes is the most common form of diabetes.

In many of these situations, the system has not been able to give perfect answers due to lack of enough data, but since focus and target of each candidate is being determined, the passage closest to what would've been an answer in an enormous corpus is being shown.

Testing results:

Out of 30 questions, answers to 22 have been acceptable. In the following results, red ones are the ones with unexpected results.

what causes gestational diabetes

" Pregnancy hormones can block insulin from doing its job. When this happens, glucose levels may increase in a pregnant woman's blood. You are at greater risk for gestational diabetes if you: Are older than 25 when you are pregnant
Have a family history of diabetes
Gave birth to a baby that weighed more than 9 pounds or had a birth defect
Have high blood pressure
Have too much amniotic fluid
Have had an unexplained miscarriage or stillbirth
Were overweight before your pregnancy
"

What is juvenile diabetes

' Until recently, the common type of diabetes in children and teens was type 1. It was called juvenile diabetes. With Type 1 diabetes, the pancreas does not make insulin. Insulin is a hormone that helps glucose,or sugar, get into your cells to give them energy. Without insulin, too much sugar stays in the blood. But now younger people are also getting type 2 diabetes. Type 2 diabetes used to be called adult-onset diabetes. But now it is becoming more common in children and teens, due to more obesity. With Type 2 diabetes, the body does not make or use insulin well. Children have a higher risk of type 2 diabetes if they are obese, have a family history of diabetes, or are not active, and do not eat well. To lower the risk of type 2 diabetes in children Have them maintain a healthy weight
Be sure they are physically active
Have them eat smaller portions of healthy foods
Limit time with the TV, computer, and video Children and teens with type 1 diabetes may need to take insulin. Type 2 diabetes may be controlled with diet and exercise. If not, patients will need to take oral diabetes medicines or insulin. '

How to cure type 1 diabetes

" because type 1 diabetes can start quickly and the symptoms can be severe, people who have just been diagnosed may need to stay in the hospital. if you have just been diagnosed with type 1 diabetes, youmay need tohave a checkup each week until you have good control over your blood sugar. your health care provider will review the results of your home blood sugar monitoring and urine testing. your provider will also look at your diary of meals, snacks, and insulin injections. it may take a few weeks to match the insulin doses to your meal and activity schedule. as the disease gets more stable, you will have fewer follow-up visits. visiting your health care provider is very important so you can monitor any long-term problems from diabetes. you are the most important person in managing your diabetes. you should know the basic steps to diabetes management: how to recognize and treat low blood sugar (hypoglycemia)
how to recognize and treat high blood sugar (hyperglycemia)
diabetes meal planning
how to give insulin
how to check blood glucose and urine ketones
how to adjust insulin and food when you exercise
how to handle sick days
where to buy diabetes supplies and how to store them insulin insulin lowers blood sugar by allowing it to leave the bloodstream and enter cells. everyone with type 1 diabetes must take insulin every day. insulin must be injected

under the skin using a syringe, insulin pen or pump. it cannot be taken by mouth because the acid in the stomach destroys insulin. insulin types differ in how fast they start to work and how long they last. the health care provider will choose the best type of insulin for you and will tell you at what time of day to use it.sometypesof insulin may be mixed together in an injection to get the best blood glucose control. other types of insulin should never be mixed. you may need insulin shots from one to four times a day. your health care provider or diabetes nurse educator will teach you how to give insulin injections. at first, a child's injections may be given by a parent or other adult. by age 14, most children can give their own injections. people with diabetes need to know how to adjust the amount of insulin they are taking: when they exercise
when they are sick
when they will be eating more or less food and calories
when they are traveling diet and exercise by testing their blood sugar level, people with type 1 diabetes learn which foods and activities raise or lower their sugar level most. this helps them adjust their insulin doses to specific meals or activities to prevent blood sugar from becoming too high or low. the american diabetes association and the american dietetic association have information for planning healthy, balanced meals. it can help to talk with a registered dietitian or nutrition counselor. regular exercise helps control the amount of sugar in the blood. it also helps burn extra calories and fat to reach a healthy weight. talk toyour health care provider before starting any exercise program. people with type 1 diabetes must take special steps before, during, and after physical activity or exercise. managing your blood sugar checking your blood sugar levelyourself and writing down the results tells you how well you are managing your diabetes. talk to your doctor and diabetes educator about how often to check. to check your blood sugar level, you use adevice called a glucose meter. usually, you prick your finger with a small needle called a lancet to get a tiny drop of blood. you place the blood on a test strip and put the strip into the meter. the meter gives you a reading that tells you the level of your blood sugar. keep a record of your blood sugar for yourself and your doctor or nurse. the numberswill help if you have problems managing your diabetes. you and your doctor should set a target goal for your blood sugar level at different times during the day. you should also plan what to do when your blood sugar is too low or high. low blood sugar is called hypoglycemia. blood sugar levels below 70 mg/dl are too low and can harm you. foot care people with diabetes are more likely than those without diabetes to have foot problems. diabetes damages the nerves. this can make you less able to feel pressure on the foot. you may not notice a foot injury until you get a severe infection. diabetes can also damage blood vessels. small sores or breaks in the skin may become deeper skin sores (ulcers). the affected limb may need to be amputated if these skin ulcers do not heal or become larger, deeper, or infected. to prevent problems with your feet: stop smoking if you smoke.
improve control of your blood sugar.
get a foot exam by your health care provider at least twice a year and learn whether you have nerve damage.
check and care for your feet every day. this is very important when you already have nerve or blood vessel damage or foot problems.
make sure you wear the right kind of shoes. ask your health care provider what is right for you. preventing complications your doctor may prescribe medicines or other treatments to reduce your chance of developing eye disease, kidney disease, and other conditions that are common in people with diabetes. these conditions are called complications of diabetes. "

What are the symptoms of type 1 diabetes

' high blood sugar these symptoms may be the first signs of type 1 diabetes, or they may occur when blood sugar is high: being very thirsty
feeling hungry
feeling tired all the time having blurry eyesight
feeling numbnessor feeling tingling in your feet
losing weight without trying
urinating more often for other people, these serious warning symptoms may be the first signs of type 1 diabetes, or they may happen when blood sugar is very high (diabetic ketoacidosis): deep, rapid breathing
dry skin and mouth
flushed face
fruity breath odor
nausea or vomiting, inability to keep down fluids
stomach pain low blood sugar low blood sugar (hypoglycemia) can develop quickly in people with diabetes who are taking insulin. symptoms usually appear when blood sugar level falls below 70 mg/dl. watch for: headache
hunger
nervousness
rapid heartbeat (palpitations)
shaking
sweating
weakness
'

How to prevent diabetes insipidus

' there is no known way to prevent congenital nephrogenic diabetes insipidus. treating the disorders that can lead to the acquired form of the condition may prevent it from developing in some cases. medications should only be used under the supervision of the health care provider. '

Tools and Technologies used

- Python 2.7
- NLTK library