Reunião 06/02/2019

O que faremos com o survey? Vai virar artigo ou vai ser parte da minha tese? Aline acha estranho em publicar um survey com artigos que ainda estão no archive.

Para a defesa da proposta até o meio de 2020 preciso submeter um artigo.

Pendências da reunião anterior

- Olhar pendências no caderno.

- aaa

SE EVOLUIR, PODEMOS FAZER REUNIAO NO FINAL DA SEMANA QUE VEM!!!

URGENTE!!! Qual dataset o artigo que faz join utiliza? Como esse dataset foi criado?

Como vou criar

Tabela de jogadores

Criamos uma tabela artificial

Para cada uma das 24k tabelas posso criar uma tabela lateral com nome de um objeto hipotético, id e chave estrangeira. Poderei criar artificialmente!

Para simular um relacionamento N-N preciso criar uma nova tabela. Nesse caso, também podemos gerar situações sofisticadas com relação ao histórico. Posso fazer perguntas: “Qual foi o primeiro time do Neymar? ” ou “em que ano o neymar jogou no Barcelona?” ou “Qual é o atributo x do jogador associado ao time y?”

Para cada uma das 24k tabelas posso gerar: 1-1, 1-n, n-n com atributo.

URGENTE!!! Para a próxima reunião, trazer exemplos de schemas e consultas dos schemas no WikiSQL? Eles querem ver se nossa ideia é muito naive...

GERAR UM TESTE MAIS SOFISTICADO DE MANIERA MANUAL PARA CHECAR SE A SITUAÇÃO MANUAL É CAPTURADA PELO TREINO NO MODELO COM SITUAÇÕES ARTIFICIAIS. CRIAR BASE DE DADOS DE VALIDAÇÃO MANUALMENTE BEM REAIS.

Pauta de hoje

1. Mudança de área
2. Criação de um novo dataset **WikiSQL2** baseado no dataset WikiSQL (<https://github.com/salesforce/WikiSQL>). O WikiSQL é um grande corpus anotado para o desenvolvimento de interfaces de linguagem natural que possui aproximadamente 50 mil pares de registros com duas colunas: a primeira é uma pergunta em inglês e a segunda é uma consulta SQL equivalente. Os registros são baseados em informações da Wikipedia. A proposta do WikiSQL2 é enriquecer o WikiSQL através da inclusão de consultas mais complexas que considerem junções entre tabelas. Atualmente cada registro do WikiSQL consulta apenas uma tabela por vez.

- ver quais são os outros datasets utilizados pelos artigos que trabalham com junção.

- 3 pessoas falam se está certo ou errado. Para cada anotação um grupo impar da um rótulo sim ou não para validar.

1.1. Avaliação de modelos já propostos para o WikiSQL no conjunto de dados do WikiSQL2.

1.2. Criação de um modelo de parser semântico chamado **NewNet** baseado no WikiSQL2.

1.3. Comparar os resultados do NewNET com os resultados dos modelos já existentes (item 1.1) sobre o dataset WikiSQL2.

1. Criação de um novo dataset **WikiSQL2BR** baseado no conteúdo.

- vou iniciar a partir da wikipedia brasil

- ou vou me propor a traduzir o dataset em inglês.

1. Criação de um modelo de parser semântico chamado **CCRNet** baseado no WikiSQLBR que converte perguntas escritas em português para queries SQL.

- dependendo de como funcionar o NewNET, é prematura pensar em uma versão para o inglês.

Próximos passos

- como pegar o dump do Wikipedia?

- Ler (muito importante) e escrever 1 parágrafo sobre cada texto e enviar para orientadores:

-- <https://medium.com/@tao.yu/spider-one-more-step-towards-natural-language-interfaces-to-databases-62298dc6df3c>

-- <https://www.salesforce.com/blog/2017/08/salesforce-research-ai-talk-to-data>

-- <https://medium.com/@jrodthoughts/this-cool-technology-brings-natural-language-processing-to-databases-58449a6664cb>

- Escrever programa que faça a leitura do WikiSQL.

- Escrever programa que gere conteúdo no formato WikiSQL.

- Ler detalhadamente o artigo: What It Takes to Achieve 100% Condition Accuracy on WikiSQL - http://aclweb.org/anthology/D18-1197.

- Escrever texto explicando o WikiSQL.