

## Recuperação de Informação na Web – 2017/1 Prof. Álvaro R. Pereira Jr.

### Trabalho Prático – Apresentação geral

As áreas de Recuperação de Informação e Web são hoje base para muitas atividades em Ciência da Computação, dentre elas atividades das áreas de aprendizado de máquina, processamento de linguagem natural, descoberta de conhecimento e mineração de dados Web.

Nesse trabalho prático vamos explorar as técnicas de RI, em especial técnicas associadas à Web, no contexto de uma área de aplicação, que é a área jurídica. Em especial, vamos explorar a área de contratos.

Queremos desenvolver uma aplicação que faça uso de conhecimento descoberto na Web para realizar algum serviço que facilite com que o público geral e/ou profissionais da área jurídica (em especial, advogados) possam realizar alguma nova atividade com valor agregado de forma automática ou semiautomática.

Por exemplo, seria possível pensar em uma aplicação que auxilia o usuário a montar um contrato, a partir de partes daquele tipo de contrato que podem ser encontradas em diferentes instâncias na Web, em diferentes fontes, no intuito de escolher as melhores cláusulas, de forma a respeitar a legislação vigente e ainda incorporar as demandas específicas daquele usuário?

Certamente, essa seria uma inovação disruptiva com alto valor agregado para o usuário final, seja ele um advogado ou um cidadão comum que visa fazer o seu contrato por conta própria. Até o momento, não se conhece aplicação similar.

Ainda que essa possível aplicação demande um bom volume de trabalho em pesquisa em algumas áreas da Ciência da Computação, nós estamos interessados em saber se conseguimos evoluir com a aplicação de uma forma ingênua, ou seja, usando das técnicas mais básicas, ou permitindo que os estudantes evoluam as suas aplicações a partir de conhecimentos de técnicas que extrapolam o programa do curso.

Para a aplicação alvo, algumas questões estão abertas, a saber:

- Com que qualidade (ou eficácia) conseguiremos coletar os diferentes tipos de contratos na Web? Ou seja, conseguiremos encontrar bons modelos de contratos? Conseguiremos encontrar muitos contratos?
- Com que qualidade conseguiremos classificar os contratos em classes (ou tipos de contratos)?
- Poderemos usar de fontes adicionais de dados na Web para enriquecer a qualidade da aplicação, de alguma forma? Por exemplo, os contratos

na Web citam leis, que podem ser coletadas e juntadas à base de dados.

- Como podemos propor cláusulas que são mais favoráveis à parte que está apresentando o contrato?
- Como avaliaremos a solução em desenvolvimento?
- Se não tivermos boa qualidade, há algum serviço intermediário que poderia ser derivado? Em outras palavras, a solução serve para alguma coisa?

## Parte 1

Entrega: até 8/6/17, quinta-feira

O que deve ser feito:

1. Informar os nomes dos dois membros da equipe
2. Apresentar uma lista com pelo menos 20 sites que serão semente da coleta
3. Apresentar um escopo de tipo de contrato que a equipe pretende focar – em um primeiro momento, vamos assumir que a sua aplicação funcionará apenas para um tipo de contrato, a escolha da equipe
4. Apresentar uma justificativa para a escolha do item 3
5. Apresentar ideias de sub-produtos que a solução em desenvolvimento poderia derivar – acima uma aplicação “muito difícil” foi proposta para servir de direção, no entanto, a equipe pode ter ideias de outras aplicações que poderiam ser derivadas.
6. Apresentar ideias de COMO a solução em foco pode ser desenvolvida – será que você tem alguma ideia especial de solução que poderia ser importante para a qualidade final da aplicação?