

5 Conclusões e trabalhos futuros

Nessa dissertação, abordamos o problema de alinhamento de esquemas complexos (com mais de uma classe). Essa operação é de suma importância para aplicações envolvendo interoperabilidade ou integração de dados, como mediadores Web, que precisam interligar dezenas de bases de dados espalhadas pela Internet (Madhavan et al. [27] estimam que existam cerca de vinte e cinco milhões de bases de dados não mapeadas no universo da internet). Para endereçar o problema, discutimos as estratégias de alinhamento existentes.

No capítulo 2, descrevemos a estrutura de um arquivo RDF/OWL e endereçamos o problema de alinhamento de esquemas, classificando as diferentes técnicas existentes e descrevendo em maiores detalhes o algoritmo proposto por Leme [20] para alinhamento de esquemas.

No capítulo 3, apresentamos o *Matchmaking*, uma infraestrutura para alinhamento de esquemas. As idéias da ferramenta foram baseadas no algoritmo de Leme [20], cuja técnica é baseada no cálculo da proximidade semântica através da utilização de funções de similaridade. Descrevemos em que pontos a ferramenta é flexível e como um desenvolvedor pode utilizá-la para testar seus próprios algoritmos, funções de similaridade e representações de dados. Apresentamos como obter informações de proveniência de tal forma que, na medida do possível, o processo não interfira no trabalho do algoritmo. Apesar de a infraestrutura utilizar esquemas descritos em ontologias OWL, são necessárias pequenas modificações em seu código para que o *Matchmaking* consiga processar outras tecnologias.

No capítulo 4, descrevemos como testamos o *Matchmaking*, implementando parte do algoritmo proposto por Leme, suas funções de similaridade e representações de dados. Também criamos uma interface gráfica que permitiu testar cada uma das funcionalidades disponíveis, realizar as operações de alinhamento de forma simplificada e verificar o histórico dos alinhamentos realizados, com seus dados de proveniência. Por último descrevemos o

MatchmakingBatch, uma ferramenta para alinhamento de esquemas em lote. Nela, é possível, a partir de uma única operação, executar diversos algoritmos de alinhamento (com diferentes valores de parâmetros) sobre dois esquemas.

A seguir, listamos alguns pontos que podem ser considerados nos trabalhos futuros:

- **Implementação de todos os passos do algoritmo.** Dada a complexidade dos terceiro e quarto passos do algoritmo de exemplo, não foi possível concluir este item. Apesar disso, seria importante conseguir automatizar todo o processo de alinhamento proposto por Leme [20]
- **Implementar novos algoritmos de alinhamento.** O *Matchmaking* foi criado para ser flexível o suficiente ao ponto de podermos implementar diversos algoritmos de alinhamento, segundo técnicas diferentes. Um trabalho futuro seria experimentar algum outro algoritmo de alinhamento proposto na literatura atual e ver como ele se adapta à ferramenta e se ela precisa prover mais alguma informação que hoje não está disponível.
- **Mediador de dados.** Um mediador de dados é uma ferramenta que agrega vários esquemas de mesmo domínio e, com a informação contida em cada um deles, consegue criar um esquema mediado que é a união (normalizada) de todas as bases agregadas. Com a infraestrutura de alinhamento pronta, é possível utilizá-la integrada a um serviço de mediação de esquemas que, ao receber um novo esquema, o compara com o mediado através do *Matchmaking* e o agrega ao ambiente de mediação.
- **Calibrador de domínio.** Os algoritmos de alinhamento são altamente parametrizados. Porém, dependendo dos valores informados, o resultado de um alinhamento pode variar enormemente. Além disso, valores utilizados anteriormente em outro domínio de aplicação e que resultaram em um alinhamento satisfatório podem não atender ao problema corrente. Um estudo interessante seria criar uma ferramenta capaz de calibrar um algoritmo de alinhamento para diferentes domínios. Ela poderia

fazer isso variando os valores até alcançar um nível, por exemplo.