

2 Interoperabilidade de Ontologias

Ontologias representam entidades importantes, também chamadas de conceitos, com suas relações taxonômicas e não-taxonômicas, e as restrições aplicadas nestas entidades de domínio. Em algumas representações, pode-se até usar axiomas para declarar as verdades que são válidas no Universo de Discurso de uma ontologia, i.e., o contexto onde uma ontologia está inserida.

Na Web Semântica, onde as informações são heterogêneas e distribuídas, o uso de ontologias, além de prover uma teoria do domínio, provê também uma representação comum para a troca das informações. Desta maneira, diferentes agentes de software que atuam neste ambiente obtêm suporte no processamento automático de informações. No entanto, para que trabalhem em conjunto, trocando automaticamente as informações representadas em ontologias, são necessários mecanismos que garantam a interoperabilidade de ontologias.

2.1. Mecanismos

Atualmente existem algumas aproximações para o processo de compatibilidade de ontologias na Web Semântica, entre elas: (1) combinação de ontologias (Noy e Musen, 1999a), (2) alinhamento de ontologias (Noy e Musen, 1999a), (3) mapeamento de ontologias (Noy e Musen, 2003), (4) integração de ontologias (Pinto et al., 1999), entre outras. Tais mecanismos são descritos a seguir:

Combinação de Ontologias – Na combinação de ontologias tem-se como resultado a versão das ontologias originais combinadas em uma ontologia única, com todos seus termos juntos e sem a definição clara de suas origens. Normalmente as ontologias originais descrevem domínios similares² ou de

² Neste trabalho, entende-se por domínios similares aqueles domínios que tratam do mesmo assunto.

sobreposição³. A Figura 2 ilustra um exemplo de combinação de ontologias, onde os dois conceitos compatíveis, **carro** da ontologia **O1** e **veículo** da ontologia **O2**, são combinados, i.e., unidos, na ontologia única **O**.

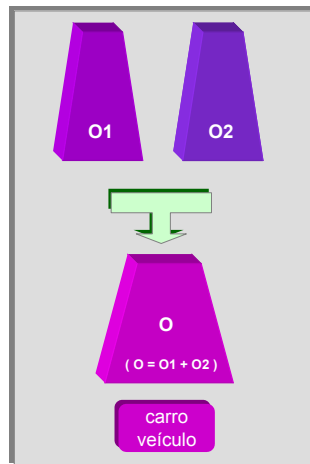


Figura 2 – Combinação de ontologias, adaptado de (Noy, 1999b)

Alinhamento de Ontologias – No alinhamento de ontologias tem-se como resultado as duas ontologias originais separadas, mas nestas são adicionadas às ligações entre seus termos equivalentes. Estas ligações permitem que as ontologias alinhadas reusen as informações umas das outras. O alinhamento normalmente é realizado quando as ontologias são de domínios complementares⁴. A Figura 3 ilustra um exemplo de alinhamento de ontologias, onde o conceito **carro** de **O1** é alinhado, i.e., ligado, ao conceito **veículo** de **O2**.

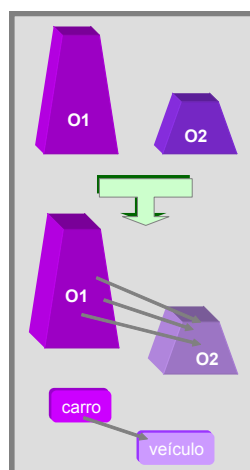


Figura 3 – Alinhamento de ontologias, adaptado de (Noy, 1999b)

³ Neste trabalho, entende-se que um domínio A qualquer é de sobreposição a um domínio B qualquer, quando A além de possuir as informações contidas em B também possui informações adicionais sobre o mesmo assunto de B.

⁴ Neste trabalho, entende-se que um domínio A qualquer é complementar a um domínio B qualquer, quando A adiciona informações a B. Normalmente, domínios complementares tratam de assuntos diferentes, mas suas partes comuns tratam de um mesmo assunto.

Mapeamento de Ontologias – No mapeamento de ontologias tem-se como resultado uma estrutura formal com expressões que ligam os termos de uma ontologia nos termos de uma outra ontologia. Este mapeamento pode ser usado para transferir instâncias de dados, esquemas de integração e de combinação, e outras tarefas similares. A Figura 4 ilustra um exemplo de mapeamento de ontologias, onde os conceitos **carro** de **O1** e **veículo** de **O2** são mapeados em expressões formais.

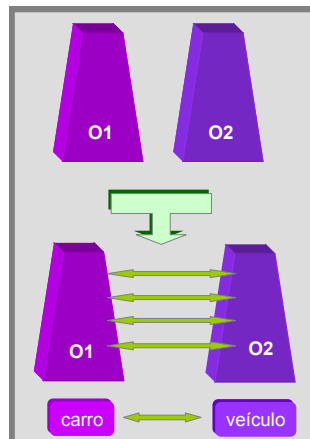


Figura 4 – Mapeamento de ontologias, adaptado de (Noy, 1999b)

Integração de Ontologias – Na integração de ontologias tem-se como resultado uma ontologia única criada pela montagem, extensão, especialização ou adaptação de outras ontologias de assuntos diferentes. Na integração de ontologias é possível identificar as regiões que foram criadas a partir das ontologias originais. A Figura 5 ilustra um exemplo de integração de ontologias, onde os conceitos **carro** de **O1**, **veículo** de **O2**, **automóvel** de **O3** e **meio de transporte terrestre** de **O4** são integrados, i.e., unidos, na ontologia única **O**.

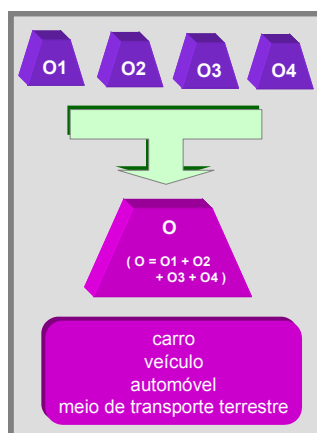


Figura 5 – Integração de ontologias

Apesar do resultado final tanto da combinação quanto da integração de ontologias ser uma ontologia única, constituída pela união dos termos das

ontologias originais, a principal diferença entre estes dois mecanismos é que no primeiro as ontologias tratam do mesmo assunto, o que não acontece necessariamente no segundo.

Neste trabalho busca-se a identificação dos termos equivalentes de aplicações complementares e o endereçamento destes termos de forma a permitir a negociação de suas informações. Por estas razões, o mecanismo de alinhamento de ontologias é o escolhido.

2.2.

Alinhamento de Ontologias

Entende-se por alinhamento de ontologias o processo a que as diversas aplicações de sistemas abertos, com suas diferentes ontologias, terão que se submeter para garantir uma representação intermediária da informação que poderá ser compartilhada entre elas. O resultado deste processo é um modelo persistente que estabelece a ligação entre as ontologias alinhadas, permitindo o compartilhamento e a reutilização de suas informações.

Para exemplificar, suponha que dois agentes de software, de dois sites específicos de aplicações complementares, desejam realizar uma negociação. Um agente refere-se a uma aplicação do domínio *Fast Food* e o outro a uma aplicação do domínio de Bebidas. Na aplicação do domínio *Fast Food* só existem bebidas não-alcoólicas, mas na do domínio de Bebidas existem tanto bebidas não-alcoólicas como alcoólicas. O processo de alinhamento deverá garantir que não haverá a inclusão de bebidas alcoólicas na negociação entre as aplicações dos domínios de *Fast Food* e Bebidas, ou seja, que só as informações comuns, interseção ilustrada na Figura 6, serão as informações compartilhadas entre elas.

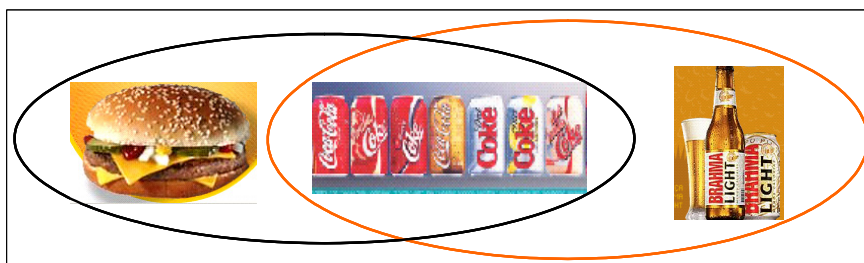


Figura 6 – Interseção de mercadorias de aplicações complementares

Para a identificação dos termos a serem alinhados em ontologias comparadas, bons algoritmos de identificação são necessários de forma a minimizarem possíveis problemas como: presença de inconsistências encontradas nas comparações dos termos de ontologias; falta de completude no alinhamento de

todos seus possíveis termos; e risco de inviabilidade da estratégia frente às inconsistências existentes. Aspectos que devem ser tratados para evitar possíveis inconsistências no processo de alinhamento são apresentados a seguir. Os demais problemas são avaliados no tópico 5.2. deste texto.

2.2.1.

Compatibilidade de Termos

Para alinhar diferentes ontologias é necessário listar as diferenças entre seus termos, perceber as similaridades entre estes, detectar tanto possíveis inconsistências quanto a falta de informação (completeza). Preocupa-se, em particular, com a identificação de:

1. Conceitos com um mesmo significado, mas rotulados com nomes diferentes;
2. Conceitos rotulados com o mesmo nome, mas com significados diferentes;
3. Diferenças na escrita dos termos das ontologias como, por exemplo: um termo no plural e outro no singular, um no feminino e outro no masculino, e em diferentes tempos verbais;
4. Propriedades com um mesmo significado, mas rotuladas com diferentes nomes;
5. Propriedades rotuladas com o mesmo nome, mas com significados diferentes;
6. Diferenças nas restrições e nas propriedades utilizadas;
7. Diferenças nas propriedades utilizadas em restrições – conceitos relacionados que são similares;
8. Diferenças nos conceitos relacionados utilizados em restrições – propriedades que são similares;
9. Diferenças no número de restrições; (diferenças nos casos onde exista uma interseção de restrições);
10. Verificação se todos os conceitos a que as propriedades se relacionam são equivalentes em ambas as ontologias comparadas;
11. Verificação se todos os conceitos que utilizam uma propriedade são consistentes em ambas as ontologias comparadas.

A realização de cada identificação numerada acima refina a qualidade do resultado obtido com o alinhamento, funcionando como um medidor de inconsistência entre as ontologias pesquisadas. Uma vez identificadas as inconsistências existentes, é preciso decidir qual ação será tomada. Como

apresentado em Nuseibeh et al. (2000), as três estratégias seguintes são possíveis: ignorar, tolerar ou resolver.

Não é desejado que inconsistências sejam simplesmente ignoradas, pois é preciso realizar o alinhamento de ontologias com confiabilidade mas, por outro lado, não é desejado resolver todas as inconsistências existentes, pois isto aumentaria a complexidade da solução e, conseqüentemente, seu tempo de execução. Por estas razões, a existência de algumas inconsistências, de acordo com um percentual pré-definido de aceitação, é tolerável.

Realizando uma série de experimentos para alinhamento e analisando seus resultados obtidos, pode se chegar, empiricamente, a uma região com os valores para um percentual de similaridade razoável. O percentual de similaridade escolhido deve ser utilizado como medida de tolerância para as inconsistências detectadas. Uma estratégia de escolha de onde o percentual de similaridade será aplicado deve ser elaborada para definir quais serão as inconsistências toleradas.

A seguir, são apresentados os requisitos e os resultados considerados satisfatórios para o alinhamento de ontologias no contexto da Web Semântica.

2.2.2.

Requisitos

O processo de alinhamento de ontologias pode ser realizado de forma automática, semi-automática, onde há a necessidade de intervenção humana em algumas etapas para a tomada de decisão, ou até mesmo manual. Hoje, a Internet possui mais de oito bilhões de páginas, segundo a ferramenta de busca Google⁵. Supondo que na Web Semântica a maioria de suas páginas terá sua própria ontologia, o alinhamento manual, ou até mesmo o semi-automático, passa a ser tedioso, sujeito a erros e difícil na escala desta Web.

Na visão da Engenharia de Software, é desejável que o processo de alinhamento seja automático e, se possível, rápido (executado em poucos segundos) e confiável (mínima margem de erro nos alinhamentos realizados). Desta forma, sistemas multi-agentes (OMG, 2000), por exemplo, têm alguns de seus requisitos de interoperabilidade satisfeitos (Haendchen et al., 2003).

O alinhamento rápido se faz necessário pela própria natureza dinâmica da Web. Entende-se que, ao ser disparado o pedido de alinhamento pelos agentes de

⁵ <http://www.google.com>

software, a resposta deve ser, se possível, imediatamente depois da execução do alinhamento. Também, intervenções durante o processo de alinhamento não são desejadas porque atrasariam este processo. Além do mais, acredita-se que usuários e agentes de software são especialistas em aplicações Web e não, necessariamente, no domínio onde estas estão inseridas. Desta forma não interviriam com informações relevantes no processo decisivo de alinhamento.

O alinhamento total é desejado, porém, desnecessário. Prefere-se manter um critério de alinhamento confiável, que pode resultar em alguns termos não alinhados, mas que, por outro lado, apresente um baixo percentual de erro na identificação dos termos comuns, com um tempo de execução razoável. Soluções complexas para o alinhamento total, que requerem elevado tempo de execução, não parecem ter sucesso neste contexto e, conseqüentemente, perdem sua justificativa de uso.

Escolhas sobre os recursos a serem utilizados para o alinhamento são decididas balanceando suas contribuições com seus impactos. Qualquer nova escolha que faça com que um dos requisitos de alinhamento na visão da Engenharia de Software não seja satisfeito é descartada.

2.2.3.

Resultados Considerados Satisfatórios

O alinhamento automático, realizado em um tempo finito de execução, é considerado como uma boa resposta, desde que se respeitando os limites pré-definidos de confiabilidade. Neste caso, os resultados de alinhamentos realizados não precisam ser armazenados, porque alinhamentos são realizados sempre que necessários. Seus resultados podem ser descartáveis e, portanto, não persistentes.

Existe um custo associado caso as informações conseguidas com o alinhamento sejam armazenadas. Este custo deve ser analisado balanceando a razão de uso do armazenamento com seus possíveis resultados obtidos. Em Doan et al. (2003), por exemplo, as informações obtidas com o mapeamento de ontologias são armazenadas para que técnicas de aprendizado possam ser aplicadas. Tais técnicas não devem resultar em melhorias para o processo de alinhamento automático das ontologias de aplicações Web. Isto ocorre porque neste processo o alinhamento das mesmas ontologias deve acontecer ocasionalmente, impedindo que a informação seja “aprendida” e,

conseqüentemente, utilizada. Além do mais, a informação precisa ser interpretada para ser “aprendida”. Esta necessidade não é nada trivial quando satisfeita de forma automática.

2.3.

Revisão da Literatura

A interoperabilidade de ontologias vem sendo estudada por diferentes pesquisadores. Naturalmente, algumas abordagens distintas têm sido exploradas. Existe, por exemplo, a abordagem de uso de **ontologias preferenciais** em que seus termos são aceitos por uma determinada comunidade (domínio); uso de dicionários, thesauri, léxicos para o casamento sintático; transformação; entre outros.

Em Bouquet et al. (2003) é proposto um algoritmo para descoberta de **mapeamentos semânticos**, cruzando classificações hierárquicas, baseado em uma nova aproximação para a coordenação semântica, i.e., alinhamento. Esta aproximação traz o problema de coordenação semântica do problema da lingüística computacional ou similaridades estruturais para o problema de dedução de relações entre conjuntos de fórmulas lógicas, que representam o significado dos conceitos dos diferentes modelos. São combinadas as informações do nível do conhecimento léxico (conhecimento sobre as palavras utilizadas), do nível do conhecimento do domínio (conhecimento sobre as relações entre os sentidos das palavras) e do nível do conhecimento estrutural (conhecimento vindo de como as palavras estão organizadas na árvore de representação) para a construção de uma nova representação do problema. Nesta, o significado de cada nó é codificado como uma fórmula lógica e os conhecimentos relevantes do domínio e das relações estruturais entre os nós são adicionados aos nós como conjuntos de axiomas. Assim, o problema de descoberta da relação semântica entre estes nós passa a ser traduzido não como um problema de mapeamento, mas sim como um simples problema de dedução lógica.

Para o trabalho com múltiplas e extensas ontologias, é apresentado em Noy e Musen (2003) o conjunto de ferramentas *PROMPT*. Dentre as ferramentas deste conjunto, citamos: *iPROMPT*, uma ferramenta interativa para **combinação de ontologias**; *AnchorPROMPT*, uma ferramenta automática baseada em grafos para **alinhamento de ontologias**; *PROMPTFactor*, uma ferramenta para **extração de**

partes de ontologias e *PROMPTDiff*, uma ferramenta para **identificação de diferenças entre duas versões da mesma ontologia**. O conjunto de ferramentas *PROMPT* e como estas estão relacionadas são ilustrados na Figura 7.

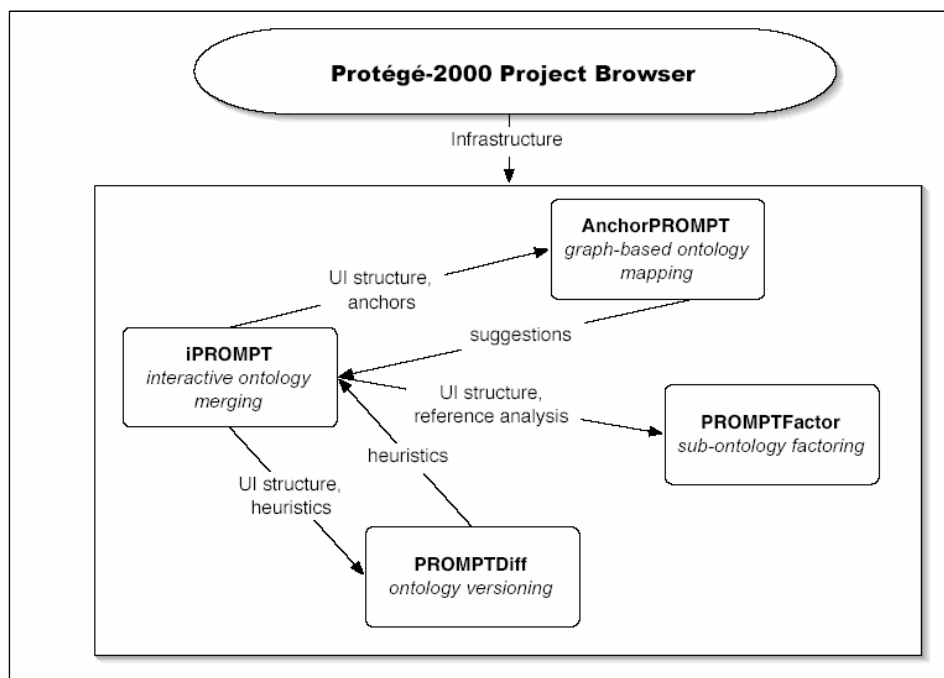


Figura 7 – O conjunto de ferramentas *PROMPT* de (Noy e Musen, 2003)

A ferramenta *iPROMPT* é uma ferramenta semi-automática de **combinação de ontologias**. Esta ferramenta guia o usuário, apresentando sugestões para os termos das ontologias que devem ser combinados, e identifica inconsistências e problemas potenciais, sugerindo estratégias para resolvê-los. Seu algoritmo faz uso tanto da informação da estrutura dos conceitos na ontologia e relações entre eles quanto da informação obtida do usuário. No entanto, as informações analisadas entre os conceitos são limitadas ao contexto local, ou seja, apenas são analisadas as informações das relações entre os conceitos ligados diretamente e aqueles que são referenciados nas restrições destas relações.

A ferramenta *Anchor_PROMPT* é uma ferramenta para o **alinhamento de ontologias** que encontra automaticamente termos semanticamente similares. Esta ferramenta tem como entrada um conjunto de âncoras, i.e., pares de termos relacionados, definidos pelo usuário ou por identificação automática de combinação lexical, e trata uma ontologia como um grafo. Neste grafo, os conceitos das ontologias são seus nós e suas relações são seus *links*, i.e., ligações. Os caminhos do sub-grafo limitado pelas âncoras são analisados e são determinados quais os conceitos que frequentemente aparecem nas mesmas

posições dos caminhos similares. Com estas análises e determinações, além do contexto local, o contexto não-local também é analisado. Além disto, utiliza na detecção de termos similares, a relação estrutural dos termos de ontologias comparadas, medidas de similaridades pré-definidas e grupos de equivalência.

A ferramenta *PROMPTFactor* é uma ferramenta para **separação de sub-ontologias, semanticamente independentes**, de ontologias extensas. Esta ferramenta é de grande auxílio para autores de ontologias que desejam criá-las reutilizando partes de ontologias extensas já criadas. Isto porque, tal ferramenta trata-se de uma possível solução para o problema de reutilização de termos de ontologias extensas sem ter que importá-las em sua totalidade. É conhecido que, ao importar uma ontologia, há a adição de todos os termos da ontologia importada na que faz uso da importação. Para as ontologias que utilizam apenas um conjunto de termos de ontologias importadas, o modelo fica desnecessariamente robusto com o excesso de informação acrescido. Este problema é agravado quando a ontologia importada é extensa.

A ferramenta *PROMPTDiff* é uma ferramenta para **identificação automática de diferenças entre duas versões da mesma ontologia**. Nesta ferramenta, a comparação não é feita apenas por comparação de textos, como tradicionalmente é realizado em comparação de versões de programas, mas também, por comparação estrutural. O *PROMPTDiff* compara as estruturas de duas versões da mesma ontologia identificando as partes que não tiveram alteração alguma, aquelas que tiveram alterações apenas em suas propriedades e aquelas que tiveram alterações nos nomes e/ou estruturas tanto de seus conceitos quanto de suas propriedades, entre outras partes.

O sistema GLUE (Doan et al., 2003) faz uso de **estratégias de aprendizagem múltiplas** para encontrar os **mapeamentos semânticos** entre duas ontologias. Neste sistema, dadas duas taxonomias e suas instâncias associadas, para cada nó, i.e., conceito de uma taxonomia, o sistema encontra o nó mais similar na outra taxonomia, dada uma medida de similaridade pré-definida. É previsto no sistema a utilização da maioria das medidas práticas de similaridades conhecidas. Trata-se de uma ferramenta de suporte automático que disponibiliza seus resultados para uma outra ferramenta qualquer de interoperabilidade semântica, não necessariamente automática, que fará uso destes resultados.

Contudo, existe uma redução significativa no esforço requerido pelo usuário que, em alguns casos, é resumido à validação, eliminando a fase de construção.

Na comunidade de Banco de Dados, o problema de **mapeamento dos diferentes esquemas** de bancos é antigo e possui soluções bem sucedidas, como o uso de conversores, mediadores e técnicas de mapeamento. Além disso, soluções específicas para ontologias já estão sendo dadas pela comunidade de banco de dados. Em (Moreira, 2003), por exemplo, é proposta uma arquitetura geral para **integração semântica de sistemas de informação** baseados na linguagem de ontologia *OWL*, onde as formas de obtenção e extração de ontologias são identificadas com ênfase em sistemas de banco de dados.

O serviço *Articulation Service* disponibilizado *on-line* em (Articulation Service, 2004) realiza o **mapeamento de duas ontologias**. A primeira ontologia é a chamada de ontologia de assunto e a segunda de ontologia objeto. O mapeamento é realizado de forma assimétrica da ontologia de assunto para a ontologia objeto.

A ferramenta de **mapeamento de ontologias baseada em léxico** disponibilizada para *download* em (Teknowledge, 2004) é uma ferramenta automática desenvolvida e acessada em SWI-Prolog (SWI-Prolog, 2004). Esta ferramenta realiza tanto o mapeamento de uma única ontologia quanto o de duas ontologias, este último de forma unidirecional e bidirecional.

O serviço *OntoMerge - Ontology Translation by Merging Ontologies* disponibilizado *on-line* em (OntoMerge, 2004) realiza a **combinação de ontologias** pela união de seus axiomas. É um serviço semi-automático que dá suporte a humanos e a agentes de software na tarefa de encontrar diferenças de notação entre duas ontologias de áreas de sobreposição.

Apesar dos trabalhos realizados para garantir mecanismos que suportem a interoperabilidade de ontologias, não foi encontrada solução alguma para o alinhamento de ontologias, prioridade deste trabalho, que esteja de acordo com os requisitos da Web Semântica, apresentados no tópico 2.2.2.. No entanto, boas idéias surgiram para elaborar uma estratégia que esteja em conformidade com tais requisitos.

O Componente para Alinhamento Taxonômico de Ontologias – CATO, resultado da implementação da estratégia elaborada, alinha automaticamente as taxonomias das ontologias de entrada, além de ser rápido e confiável. O CATO é

automático porque não é permitida a intervenção do usuário para tomada de decisão no alinhamento; Rápido porque prioriza o alinhamento de conceitos das ontologias comparadas. Quando o alinhamento é disparado pelos pedidos de agentes de software, espera-se alguma resposta em um tempo finito, se possível, imediatamente depois de sua execução; Confiável porque o alinhamento é tratado em etapas, onde cada etapa possui condições a serem satisfeitas que garantam um baixo percentual de erros.

O CATO é uma aproximação de algumas soluções da Ciência da Computação para comparação lexical e estrutural, e uso de medidas de similaridades para os ajustes finos de seus resultados. Estas soluções, apresentadas a seguir, são customizadas para o problema de alinhamento de ontologias.