2 Fundamentos e estado da arte

Este capítulo descreve os principais fundamentos usados neste trabalho e os principais trabalhos relacionados que foram identificados.

2.1 Fundamentos

Os três fundamentos tratados nesta seção serviram como base para o desenvolvimento deste trabalho. Web Semântica e *Linked Open Data*, essencial para parte do trabalho que trata do compartilhamento de dados na *web* entre diferentes sistemas de informação; Sistemas de frames e slots, usado na parte do trabalho que define os estereótipos de esquemas narrativos e *UFO* (*Unified Foundational Ontology*) que descreve parte da ontologia de fundamentação usada como referência na construção de uma ontologia de futebol. Na seção dedicada a *UFO* é explicado o que é uma ontologia, o que é uma ontologia de fundamentação e qual o seu papel no processo de modelagem conceitual

2.1.1 Web Semântica e Linked Open Data

Em [Berners-Lee et al, 2001], Tim Berners-Lee, James Hendler e Ora Lassila apresentaram pela primeira vez a ideia da Web Semântica em um artigo publicado na revista *Scientific American*.

A grande oportunidade apontada no trabalho está relacionada ao fato de que a maior parte do conteúdo disponível na Web hoje é desenvolvido para humanos lerem, e não para que programas de computador possam manipular o significado destes conteúdos. Essa característica, de certa forma, limita a capacidade que os seres humanos tem de realizar tarefas pela Web com auxílio de computadores.

Uma das principais evoluções propostas para Web Semântica é a possibilidade de expressar significado para conteúdos disponíveis em páginas da Web. A arquitetura proposta cria um ambiente onde agentes de software trafegam entre páginas para facilmente realizar tarefas sofisticadas para os usuários.

A Web Semântica não é uma Web separada mas uma extensão da atual, em que a informação tem significado bem definido, permitindo melhor que computadores e pessoas trabalhem em cooperação. Até hoje, a Web se desenvolveu mais rapidamente como um meio de documentos para as pessoas em vez de dados e informações que podem ser processados automaticamente. A Web Semântica visa compensar isso.

A representação do conhecimento é uma questão importante a ser trabalhada. Um dos desafios da Web Semântica é fornecer uma linguagem que expresse dados e regras para o raciocínio sobre os dados. Essa linguagem também deve permitir que as regras de qualquer sistema de representação do conhecimento existente possam ser exportadas para a web. E isso ainda não é suficiente. Supondo que dois bancos de dados podem usar identificadores diferentes para o que é, de fato, o mesmo conceito. Um programa desenvolvido para comparar ou combinar informações entre esses dois bancos de dados deve ser capaz de identificar que dois termos estão sendo usados para representar a mesma coisa. Idealmente, o programa deve ter uma maneira de descobrir esses significados comuns para qualquer banco de dados que encontre. A solução para este problema é fornecida por um componente básico da Web Semântica, as coleções de informações, chamadas ontologias.

Na filosofia, uma ontologia é uma teoria sobre a natureza da existência, dos tipos de coisas existentes; ontologia como disciplina estuda tais teorias.

Pesquisadores de inteligência artificial e Web tem adaptado o termo para o seu jargão próprio, e para eles uma ontologia é um documento ou arquivo que define formalmente as relações entre os termos.

As ontologias podem melhorar o funcionamento da Web, em muitos aspectos. Por exemplo, elas podem ser usadas de uma forma simples para melhorar a precisão das pesquisas na Web. O programa de pesquisa pode procurar com precisão apenas as páginas que se referem a um conceito específico em vez de todas as páginas usando palavras-chave que podem gerar resultados ambíguos.

A Web Semântica promove sinergia: mesmo agentes de software que não foram expressamente concebidos para trabalharem juntos podem transferir dados

entre si quando os dados estão semanticamente representados. Os agentes consumidores e produtores podem chegar a um entendimento comum através do compartilhamento de ontologias que proveem o vocabulário necessário para discussão. Os agentes podem desenvolver novas capacidades de raciocínio quando eles descobrem novas ontologias.

No que tange a evolução do conhecimento, pode-se dizer que, a Web Semântica adequadamente projetada pode ajudar o conhecimento como um todo.

A nomeação de todos os conceitos simplesmente por uma *URI (Uniform Resource Identifier)*, permite que qualquer agente expresse novos conceitos com um esforço mínimo. A linguagem lógica unificadora da Web Semântica permitirá que esses conceitos sejam progressivamente ligados em uma Web universal. Esta estrutura abrirá o conhecimento para análises de significados por agentes de software, provendo uma nova classe de ferramentas com as quais poderemos viver, trabalhar e aprender juntos.

[Allemang 2010] descreve como é uma "empresa de dados interligados" (*Linked Data Enterprise*) onde tecnologias da Web Semântica são usadas para abordar questões fundamentais que impedem as empresas de atingir a agilidade que necessitam. A agilidade ou a capacidade de uma organização para lidar com mudanças organizacionais tornou-se um fator fundamental para manter competitividade.

Quando bancos de dados relacionais ocuparam um lugar central na gestão da informação, foi possível pensar em banco de dados como um repositório onde se iria encontrar todas as informações sobre o negócio. O banco de dados era um "local" onde as perguntas poderiam ser respondidas.

No cenário atual de informação distribuída esta metáfora não se sustenta. Esperamos que as informações venham até nós, para estar disponível em nossos *desktops*, em nossos telefones, para viajar com a gente sobre a terra e no ar. A Web nos acostumou a ter uma ampla variedade de fontes em nossas mãos. Motores de busca como Google e Yahoo! nos permitem escolher dentre diversas fontes de informação, cada uma lutando pela atenção de nossos olhos. Já é possível notar que os dias dos "repositórios únicos" de informação se acabaram. Agora esperamos uma rede de informações interligadas.

A arquitetura de uma "empresa de dados interligados" deve servir como uma descrição de todos os dados de propriedade de uma empresa; o papel que eles desempenham no processo de negócio, a quem eles pertencem e como são mantidos e geridos. Acima de tudo, essa arquitetura deve prever que os ativos de informação de uma empresa estão em permanente mudança.

As principais questões apresentadas como causas da dificuldade que as empresas tem em gerenciar suas informações de forma ágil são:

1) Compromisso com dados legados.

Em todos os níveis da empresa, os dados foram organizados de uma maneira particular para um propósito particular.

2) Compromisso com o processo de trabalho legado

Paralelamente, aos dados legados existe o processo de trabalho legado. A inovação é difícil, e a empresa tem de continuar a fazer dinheiro, mesmo que de um modo ultrapassado.

3) Problema de indexação de grandes volumes de documentos

Estruturas de dados e ferramentas focadas em criação de documentos mas não em indexação de documentos resultam em grandes conjuntos de documentos indiferenciados. A quantidade de trabalho para indexar todos os documentos pode ser muito grande.

Existem algumas soluções tecnológicas que visam resolver os problemas apresentados. Veja alguns exemplos:

1) Repositórios de metadados

Um repositório de metadados é uma abordagem tecnológica específica para gerenciamento de dados corporativos, em que um modelo comum de metadados é definido em resposta a algumas necessidades de negócio. O metadado é usado como uma espécie de interlíngua entre as fontes de dados. Sistemas de repositório de metadados tipicamente incluem ferramentas de mapeamento de fontes de dados legados (normalmente, bancos de dados relacionais) para o repositório de meta-

dados. Isso ajuda a organização a manter coerência e interoperabilidade entre suas fontes de dados.

2) Vocabulários controlados

A ideia de um vocabulário controlado é mais velha do que a computação moderna. Um vocabulário é o conjunto de termos que são usados por uma organização para se referir aos elementos do seu negócio. Em um vocabulário controlado os termos devem ser selecionados e gerenciados por um indivíduo ou grupo dentro da organização.

3) Processamento de Linguagem Natural

Para grandes conjuntos de dados não estruturados é atraente ter um computador para processar texto automaticamente, e determinar algo sobre seu conteúdo. A forma mais simples de Processamento de Linguagem Natural aplicada aos dados de empresas é chamada extração de conceitos.

Sistemas de processamento de linguagem natural aplicados à pesquisa de documentos obtiveram um sucesso limitado. Por um lado, eles têm de competir com os métodos estatísticos para indexação de documentos, tais como aqueles que se tornaram populares pelo Google (que pode processar um volume muito grande de documentos), enquanto por outro lado eles têm de competir com especialistas do domínio (que podem processar documentos com precisão). Ao invés de encontrar um ponto ótimo entre esses extremos, soluções de processamento de linguagem natural deste tipo não têm dominado a arena da integração de informações empresariais.

Uma “empresa de dados interligados” é uma organização em que o ato de criação da informação está intimamente associada com o ato de compartilhar informações. A partilha de dados é tão importante como a produção do mesmo.

Em uma “empresa de dados interligados” indivíduos e grupos continuam a produzir e consumir informação de formas que são específicas para suas próprias necessidades de negócios, mas produzem a informação de uma forma que a mesma possa ser conectada a outros aspectos da empresa.

As tecnologias que foram aplicadas para a integração de dados empresariais não são vistas como concorrentes, nem como tecnologias fracassadas, mas sim como ferramentas para construção da infraestrutura de uma “empresa de dados interligados”.

A indústria editorial está num ponto de virada importante - o advento da difusão de informação por meio eletrônico tornou a produção de cópia impressa menos atraente do que foi no passado.

Um valor que um editor pode adicionar é uma melhor organização do material que ele publica. Este tipo de valor agregado tem sido parte da indústria editorial há décadas, mas agora está vindo à tona, a medida que a importância da organização da informação nas organizações se torna mais evidente.

Nos dias de hoje as empresas precisam ser mais ágeis do que nunca e essa agilidade requer uma nova maneira de trabalhar. Uma maneira que permita o envolvimento de um grande número de profissionais que lidam com as informações sem que seja necessário um longo tempo de desenvolvimento de sistemas a cada vez que uma nova informação é necessária. A integração de informações se tornou um “gargalo“ para integração de processos entre diferentes áreas de uma organização. Desta forma, pode-se dizer que a empresa ágil é a “empresa de dados interligados”.

Vocabulários controlados podem ser a chave para resolução desta situação. Usados de forma ruim podem se tornar apenas mais um conjunto de regras sem sentido a serem seguidas. Usado corretamente, um vocabulário controlado pode mediar a comunicação entre partes fornecendo uma referencia comum para descrição de dados. Nesse sentido eles podem ajudar na criação de uma “empresa de dados interligados” focada no compartilhamento de informações. A ideia de uma “empresa de dados interligados” é um ideal mas várias empresas estão aplicando na prática boa parte desses conceitos. Elas criam conjuntos de dados com foco em reuso permitindo uma que uma grande variedade de profissionais contribua com seus ativos de dados em geral.

2.1.2 Sistema de frames e slots

Em [Steven T. Rosenberg 1977] é apresentado uma das primeiras abordagens para o processamento de material textual. Nele, é discutido um modelo para mapear texto em uma estrutura de dados baseada em *frames*.

Segundo esta proposta, as funções essenciais de um processador de texto podem ser divididas em duas operações:

1) Localização de um contexto anterior, chamado de tema, num banco de dados de histórias, onde o novo conhecimento pode ser armazenado. Essa função pode ser chamada de Processo de Vinculação.

2) Mapeamento de novas informações de uma sentença para o contexto localizado no Processo de Vinculação. Para isto, foi assumido que cada nova sentença de um texto bem escrito está vinculada a algum tema.

O sistema de *frames* é organizado numa estrutura de árvore, onde informações genéricas ficam nos níveis superiores da árvore enquanto *frames* específicos que distinguem conhecimento ficam nos níveis inferiores próximos às folhas. O conhecimento genérico, incluindo informações de natureza processual, é herdado automaticamente. Cada *frame* consiste em um conjunto de *slots*. Os *slots* podem conter valores e também requisitos ou restrições sobre estes valores que são verificados com o uso de regras. Um *slot* é posteriormente identificado por meio de chaves associadas.

A hierarquia dos *frames* foi usada para definir um conjunto restrito de formas como um *frame* pode se referir a outro. As formas são: Referência direta, Referência Genérica, Referência Contextual, Referência entre *frames* e Valores padrão.

1) Referência Direta é o uso do nome de um *frame* para evocá-lo diretamente.

2) Referência Genérica é quando dois *frames* compartilham um caminho na árvore de *frames* a partir do nó mais alto e em algum ponto divergem.

3) Referência Contextual envolve o uso de um termo geral em um contexto onde o referente pode ser exclusivamente não ambíguo. Por exemplo: (S1) O relatório discute o recente problema de conflito na fronteira ; (S2) O incidente não foi considerado importante. O uso da frase "O incidente" em (S2) é uma referência não ambígua a "conflito na fronteira" em (S1).

4) Referência entre *frames* utiliza os espaços vazios de um *frame*. Por exemplo: (S3) John atirou em sua esposa; (S4) A arma era automática de calibre quarenta e cinco. A segunda frase especifica o instrumento exigido pela ação da primeira frase.

5) Valores Padrão. Foi proposta uma regra de discurso segundo a qual, se não houver ligação explícita com algum tema anterior, a sentença atual irá tratar o mesmo tema da sentença anterior.

Uma vez que uma potencial ligação é encontrada, novas informações devem ser mapeadas para o contexto indicado. Duas das maneiras em que uma frase está relacionada a um tema existente são:

A) Instanciando um *frame* que descreve um tema. Instanciar um *frame* envolve substituir os valores reais que constam nas sentenças no lugar dos valores padrão de *slots* de um *frame* tema. *Slots* não instanciados criam expectativas. Todas as expectativas associadas a um tema tornam-se candidatas a instanciação pelo conhecimento de uma nova sentença ligada ao tema.

B) Aumentando expectativas. Exemplo: (S5) John matou sua esposa; (S6) Ele usou um instrumento contundente. Embora a sentença não tenha sido instanciada, agora sabemos que a característica semântica do instrumento é que ele é contundente. Isso pode ser codificado como uma restrição nos valores para o *Slot* do instrumento. A expectativa foi reforçada por este novo requisito que qualquer novo valor deve atender.

Oito artigos de primeira página do New York Times foram analisados para contar a frequência em que cada vínculo temático ocorre.

2.1.3 UFO (*Unified Foundational Ontology*)

[Guizzardi et al, 2009] enfatiza a diferença entre os sentidos do termo ontologia na computação. Por um lado, pela comunidade de Modelagem Conceitual, o termo tem sido usado de acordo com sua definição em Filosofia: Sistema de categorias formais independente de domínio e filosoficamente bem fundamentado, que pode ser usado para enunciar modelos da realidade específicos de domínio. Por outro lado, pelas comunidades de Inteligência Artificial, Engenharia de Software e Web Semântica o termo é usado como um artefato concreto de engenharia projetado para um propósito específico sem dar muita atenção para as questões de fundamentação.

Ontologias de Fundamentação (*Foundational Ontologies*) são sistemas de categorias filosoficamente bem fundamentados e independentes de domínio que têm sido utilizados com sucesso para melhorar a qualidade de linguagens de modelagem e modelos conceituais.

Em [Guizzardi et al, 2009] foi apresentada uma ontologia formal filosófica e cognitivamente bem fundamentada, denominada UFO (*Unified Foundational Ontology*), inicialmente proposta em [Guizzardi et al, 2004], que foi desenvolvida com o propósito específico de servir como uma base para linguagens de modelagem conceitual.

Existe um interesse crescente no uso de ontologias de fundamentação (também conhecidas como ontologias de alto nível ou ontologias de nível superior) para três propósitos principais: 1) Validação de modelos conceituais, 2) Criação de diretrizes para uso dos mesmos, 3) Prover semântica de mundo real para construtos modelados.

A UFO começou como uma unificação do GFO (*Generalized Formalized Ontology*) [Heller et al, 2004] e de uma ontologia de alto nível chamada OntoClean[Guarino et al, 2002]. Foi apresentado um *framework* de base ontológica geral, que pode ser usado para avaliar sistematicamente a adequação de uma linguagem de modelagem para modelar fenômenos em um determinado domínio.

As principais categorias apresentadas foram: *Object-Object Universal* e *Moment-Moment Universal*. O cerne desta ontologia exemplifica o *Aristotelian onto-*

logical square, também chamado de *Four-Category Ontology* [Lowe 2006] que contempla as categorias apresentadas.

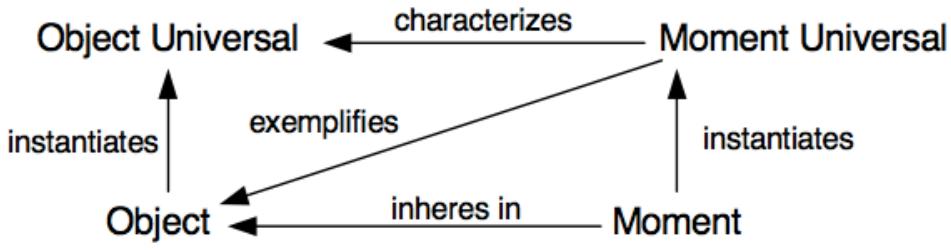

Figura 1 – Quadrado de Aristóteles. Figura extraída de [Guizzardi et al, 2009].

Uma distinção fundamental dessa ontologia é entre as categorias de Indivíduo (Particular) e Universal (Tipo). Indivíduos são entidades que existem na realidade, possuindo uma identidade única. Universais, por sua vez, são padrões de características que podem ser instanciados em indivíduos diferentes. Do ponto de vista metafísico, estas categorias permitem a construção de uma ontologia parcimoniosa, com base na noção primitiva e formalmente definida de dependência existencial: Temos que um indivíduo *x* é existencialmente dependente de outro indivíduo *y*, se e somente se, por uma questão de necessidade, *y* deve existir sempre que *x* existe. Dependência existencial é uma relação modalmente constante, isto é, se *x* é dependente *y*, esta relação entre estes dois elementos específicos ocorre em todos os *mundos* possíveis em que *x* existe.

A palavra *Moment* é derivada de *Momente* em alemão de acordo com textos de E. Husserl e denota, em termos gerais, o que às vezes é chamado de tropo, indivíduo abstrato ou instanciação de uma propriedade. Assim, no âmbito do presente trabalho, o termo não tem qualquer relação com a noção de instante de tempo em linguagem coloquial. Exemplos típicos de momentos são: uma cor, uma conexão, uma carga elétrica, um compromisso social. Uma característica importante que caracteriza todos os *Moments* é que eles só podem existir em outros elementos (por exemplo, a carga elétrica só pode existir em algum condutor). Para colocar tecnicamente, podemos dizer que *Moments* são existencialmente dependentes de outros elementos. Dependência existencial também pode ser usada para diferenciar *Moments* intrínsecos e relacionais: *moments* intrínsecos são dependentes de um único indivíduo (por exemplo, uma cor, uma dor de cabeça, uma temperatura); *moments* relacionais dependem de uma pluralidade de indivíduos (por exemplo,

um emprego, um tratamento médico, um casamento). Um tipo especial de relação de dependência existencial entre um *moment x* e um indivíduo *y*, do qual *x* depende é a relação de inerência (i). Desta forma, para que um indivíduo *x* seja um *moment* de outro indivíduo *y*, a relação i (*x*, *y*) deve existir entre os dois. Por exemplo, a inerência "liga" o seu sorriso ao seu rosto, a carga de um condutor específico ao condutor em si. Nesse ponto, admitimos que *moments* podem ser inerentes a outros *moments*. Exemplos incluem a extensão individualizada do tempo, ou a gravidade de um sintoma particular. A regressão ao infinito na cadeia de inerência é impedida pelo fato de que há indivíduos que não são inerentes a outros indivíduos, chamados objetos.

Objetos são elementos que possuem (diretamente) qualidades espaço-temporais. Exemplos de objetos incluem entidades comuns do cotidiano, como um indivíduo, um cão, uma casa, um martelo, um carro, Alan Turing e os Rolling Stones. Ao contrário dos *Moments*, os objetos não são inerentes a qualquer coisa e, como consequência, eles gozam de um maior grau de independência.

Para completar o quadrado aristotélico, consideramos aqui as categorias *Object Universal* e *Moment Universal*. Essas categorias são usadas na relação de classificação entre indivíduos e tipos. *Object Universals* classificam *Objects* enquanto *Moment Universals* classificam *Moments*.

A UFO é uma ontologia composta por três grandes fragmentos. UFO-A (cerne da ontologia UFO) foi descrita como uma ontologia de objetos que utiliza as categorias do Quadrado de Aristóteles descritas anteriormente. Os outros dois fragmentos são denominados UFO-B e UFO-C, inicialmente propostos em [Guizzardi et al, 2005]. UFO-B é uma ontologia de eventos. UFO-C fundamenta-se em UFO-A e UFO-B para sistematizar conceitos sociais.

UFO-A sistematiza conceitos como, por exemplo, tipos e estruturas taxonômicas [Guizzardi et al., 2004], relações todo-parte [Guizzardi 2007], propriedades intrínsecas e espaços de valores de atributos [Guizzardi & Masolo & Borgo, 2006], propriedades relacionais [Guizzardi & Wagner, 2008], entre outros. Esse fragmento constitui uma teoria estável, formalmente caracterizada com o aparato de uma lógica modal de alta expressividade e possuindo forte suporte empírico promovido por experimentos em psicologia cognitiva [Guizzardi 2005].

Uma distinção importante contemplada na UFO-A é entre *Substantial* e *Mode*. Substanciais (*substantial*) são indivíduos existencialmente independentes. Exemplos incluem objetos mesoscópicos do senso comum, tais como uma pessoa, um cachorro, uma casa, Tom Jobim e Os Beatles. A palavra Modo (*Mode*), em contraste, denota a instanciação de uma propriedade. Um modo é um indivíduo que só pode existir em outros indivíduos e é dito ser inerente a esses indivíduos. Exemplos típicos de modos são uma cor, uma carga elétrica, um sintoma etc. Um importante traço que caracteriza todos os Modos é o fato deles só poderem existir em outros indivíduos. Por exemplo, uma carga elétrica só pode existir em algum condutor. Com base nos conceitos de *Moment* e *Object* descritos podemos dizer que modos são *moments* intrínsecos.

Relações são entidades que aglutinam outras entidades. Na literatura de Filosofia, duas categorias amplas de relações são tipicamente consideradas, a saber relações formais e materiais. Relações formais acontecem entre duas ou mais entidades diretamente, sem nenhum outro indivíduo intervindo. Em princípio, a categoria de relações formais inclui aquelas relações que formam a super estrutura matemática do arcabouço da UFO, incluindo *dependência existencial, parte-de, subconjunto-de, instanciação*, dentre outras [Guizzardi et al, 2005]. Relações materiais, por outro lado, possuem estrutura material por si próprias e incluem exemplos como trabalhar em, estar matriculado em ou estar conectado a. Enquanto, por exemplo, a relação formal entre Paulo e seu conhecimento *x* de Grego acontece diretamente e tão logo Paulo e *x* existam, para que aconteça uma relação material *ser tratado em* entre Paulo e uma unidade médica U-M, uma outra entidade precisa existir para mediar Paulo e U-M, neste caso um tratamento.

Finalmente, foi considerada a noção de situação. Situações são entidades complexas constituídas possivelmente por vários objetos (incluindo outras situações), sendo tratadas aqui como um sinônimo para o que é chamado na literatura de estado de coisas (*state of affairs*), ou seja, uma porção da realidade que pode ser compreendida como um todo.

UFO-B é uma ontologia de eventos que diferencia explicitamente Eventos e Objetos. Eventos (ou ocorrências) são indivíduos compostos de partes temporais. Eles acontecem no tempo no sentido de se estenderem no tempo acumulando par-

tes temporais. São exemplos de eventos: uma conversa, uma partida de futebol, a execução de uma sinfonia e um processo de negócio. Eventos não podem sofrer mudanças no tempo no sentido genuíno, uma vez que nenhuma de suas partes temporais mantém sua identidade ao longo do tempo.

Eventos são possíveis transformações de uma situação para outra na realidade, i.e., eles podem alterar o estado de coisas da realidade de um (pré)estado para outro (pós-estado). Eventos são entidades ontologicamente dependentes no sentido de, para existirem, dependem existencialmente de seus participantes. Seja o evento *e*: o ataque de Brutus a César. Nesse evento, há a participação de César, Brutus e da faca usada no ataque. Neste caso, *e* é composto da participação individual de cada uma dessas entidades. Cada uma dessas participações é por si própria um evento que pode ser complexo ou atômico, mas que existencialmente depende de um único substancial. Em UFO-B, ser atômico e ser instantâneo são noções ortogonais, i.e., participações atômicas podem se estender no tempo, bem como eventos instantâneos podem ser compostos de múltiplas participações (instantâneas).

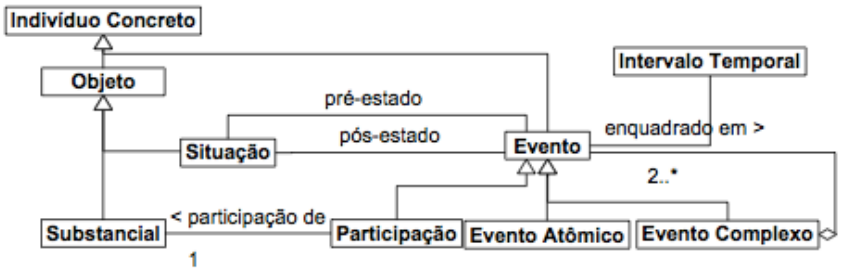

Figura 2 - Fragmento da UFO-B. Objetos e Eventos. Figura extraída de [Guizzardi et al, 2008]

Em suma, o modelo da Figura 3 mostra essas duas perspectivas sob as quais eventos podem ser analisados, a saber: como entidades que se estendem no tempo com certas estruturas mereológicas (i.e., eventos simples ou complexos), e como entidades ontologicamente dependentes que podem envolver um número de participações individuais.

Este modelo permite uma diversidade de estruturas temporais, tais como tempo linear, ramificado, paralelo e circular. Para o caso de estruturas ordenadas, consideram-se as ditas Relações entre intervalos de Allen [Allen 1983] a partir das quais as correspondentes relações entre eventos podem ser derivadas.

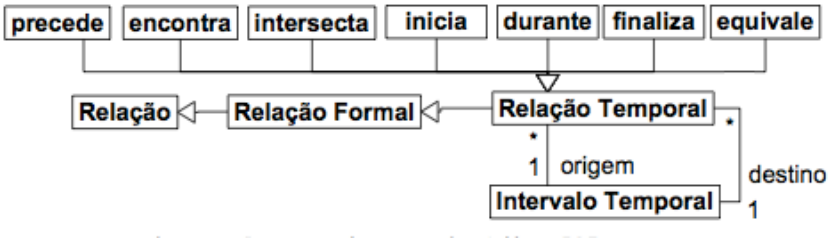

Figura 3 - Fragmento da UFO-B: Relações de Allen. Figura extraída de [Guizzardi et al, 2008].

A UFO-C é uma ontologia de entidades sociais (tanto objetos quanto eventos), construída sobre UFO-A e UFO-B.

Foi destacada a distinção entre agentes e substanciais inanimados. Agentes podem ser físicos (p.ex., uma pessoa) ou sociais (p.ex., uma organização ou sociedade). Substanciais inanimados também podem ser físicos (p.ex., um livro, um carro ou uma árvore) ou sociais (p.ex., dinheiro, linguagem e normas). Uma descrição normativa (*normative description*) [Bottazzi et al, 2008] é um tipo de substancial inanimado social que define uma ou mais regras/normas reconhecidas por, pelo menos, um agente social e que pode definir entidades sociais como universais (p.ex., tipos de comprometimentos sociais), outros objetos (a coroa do rei da Espanha) e papéis sociais, tais como presidente, ou pedestre.

Agentes são substanciais que podem possuir tipos especiais de modos chamados de modos intencionais (intentional modes). Todo modo intencional tem um tipo – p.ex, crença (*belief*), desejo (*desire*), intenção (*intention*) – e um conteúdo proposicional, sendo este último uma representação abstrata de uma classe de situações referenciadas por esse modo intencional.

Ações são eventos intencionais, i.e., eventos que instanciam um plano (ação universal) com o propósito específico de satisfazer (o conteúdo proposicional de) alguma intenção.

Atos comunicativos podem ser usados para criar modos sociais (*social modes*). Nessa visão, a linguagem não somente representa a realidade mas também cria parte dela [Searle 2000].

Comprometimentos (internos ou sociais) podem ser cumpridos (*fulfilled*) ou não cumpridos (*unfulfilled*). Comprometimentos não cumpridos podem estar pendentes (*pending*), desmarcados (*dismissed*) ou quebrados (*broken*).

Comprometimentos internos ou intenções (*intentions*) fazem com que o agente realize ações. Assim, seguindo [Conte et al, 1995], tem-se que comprometimentos sociais necessariamente causam a criação de comprometimentos internos, i.e., se Pedro promete trazer um livro amanhã para João, além do comprometimento com João, ele também tem a intenção (comprometimento interno) de trazer o livro amanhã.

Para demonstrar como a utilização dos conceitos da UFO pode ajudar na construção de modelos conceituais será apresentado a seguir um caso de real de utilização da UFO.

Em [Guizzardi et al, 2008] foi apresentado um caso de utilização da UFO para avaliar e reprojetar uma ontologia de processo de software desenvolvida no Projeto ODE (*Ontology-based software Development Environment*) [Falbo 2003]. Para facilitar o entendimento do trabalho apresentado, algumas partes do mesmo serão reescritas a seguir.

Em [Falbo et al, 2005], Falbo e Bertollo apresentam uma Ontologia de Processo de Software que foi desenvolvida para estabelecer uma conceituação comum para organizações de software falarem sobre processos de software. Essa ontologia é usada como base para o desenvolvimento de uma infraestrutura de processo para ODE (*Ontology-based software Development Environment*) [Falbo 2003], um Ambiente de Desenvolvimento de Software Centrado em Processo. A Figura 3 mostra um fragmento dessa ontologia.

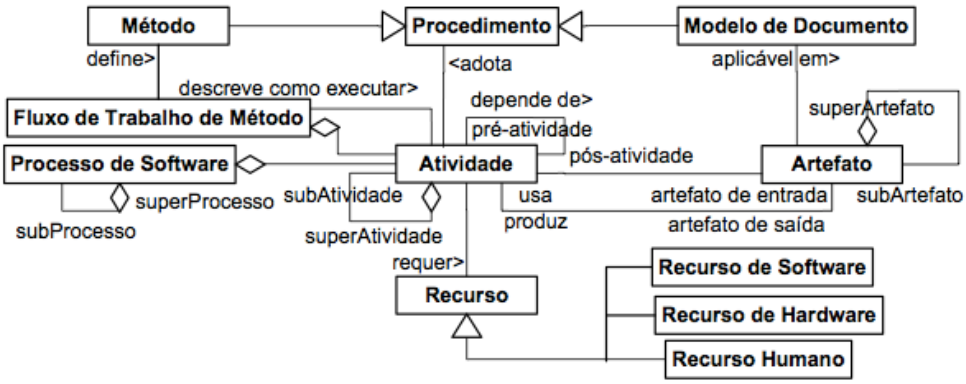

Figura 4 - Fragmento da Ontologia de Processo de Software de ODE. Extraída de [Guizzardi et al, 2008].

As Figuras 4 e 5 representam fragmentos da ontologia de processo de software revisados, obtidos pela interpretação dos conceitos da ontologia original em termos de UFO-C. Nesses modelos, os conceitos da UFO são mostrados em cinza. Nessa interpretação, torna-se claro que a ontologia de processos do ODE funde as noções de universal de ação (*action universal*) e ocorrência de ação (*action occurrence*). Fez-se, portanto, a separação desses conceitos, introduzindo-se os termos Ocorrência de Atividade e Ocorrência de Processo de Software para denotar ações particulares que ocorrem em intervalos de tempo específicos.

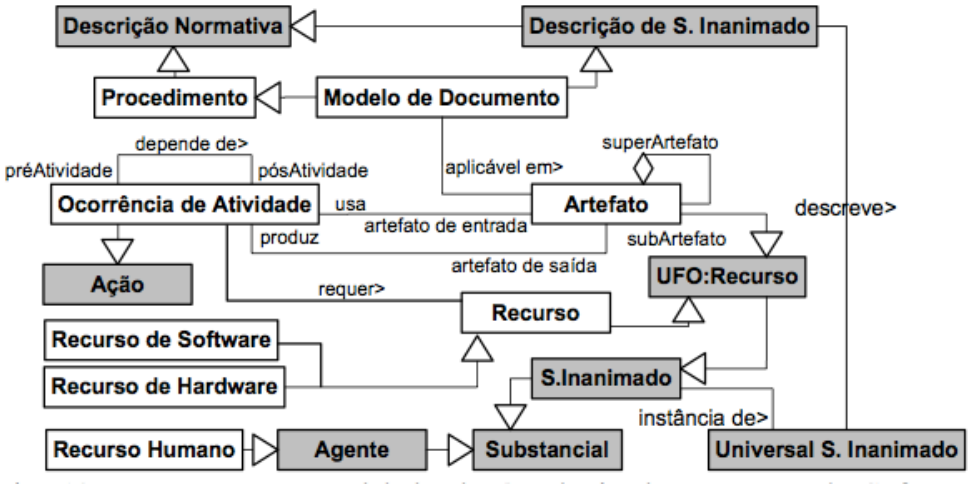

Figura 5 - Fragmento Remodelado da Ontologia de Processo de Software (Substanciais). Extraída de [Guizzardi et al, 2008].

Para eliminar a ambiguidade entre ocorrência de um sub-processo (i.e., um processo que é parte de outro processo) e ocorrência de uma atividade complexa que existia na ontologia do domínio original foi assumido que uma ocorrência de

processo de software é a raiz de um reticulado de composição, i.e, uma ocorrência de processo de software é uma ação complexa que não é parte de nenhum outro evento complexo. Uma consequência dessa definição é que não é possível haver sub-processos e, portanto, a relação *todo-parte* reflexiva entre ocorrências de processo de software foi removida do modelo.

Na ontologia original, um artefato é um tipo de Substancial Inanimado. A relação de sub-artefato entre artefatos é, portanto, governada pelos axiomas mereológicos definidos para relações todo-parte entre substanciais definidos em [Guizzardi 2005].

A noção de recurso na ontologia do ODE pode ser mapeada para a noção de substancial na UFO-A e a relação *requer* subjuga diferentes tipos de participação de uma ocorrência de atividade. Acredita-se que essa questão precisa ser elaborada nessa ontologia, de modo a fazer justiça à distinção entre contribuições de ação e participação de recursos em UFO-C. Em particular, um recurso humano (um Agente em UFO-C) não pode ser usado, modificado, criado ou eliminado por uma ocorrência de atividade.

Recursos de hardware e de software são tipos de substanciais inanimados e, portanto, seus tipos de participação devem corresponder aos tipos de participação de recurso definidos na UFO-C.

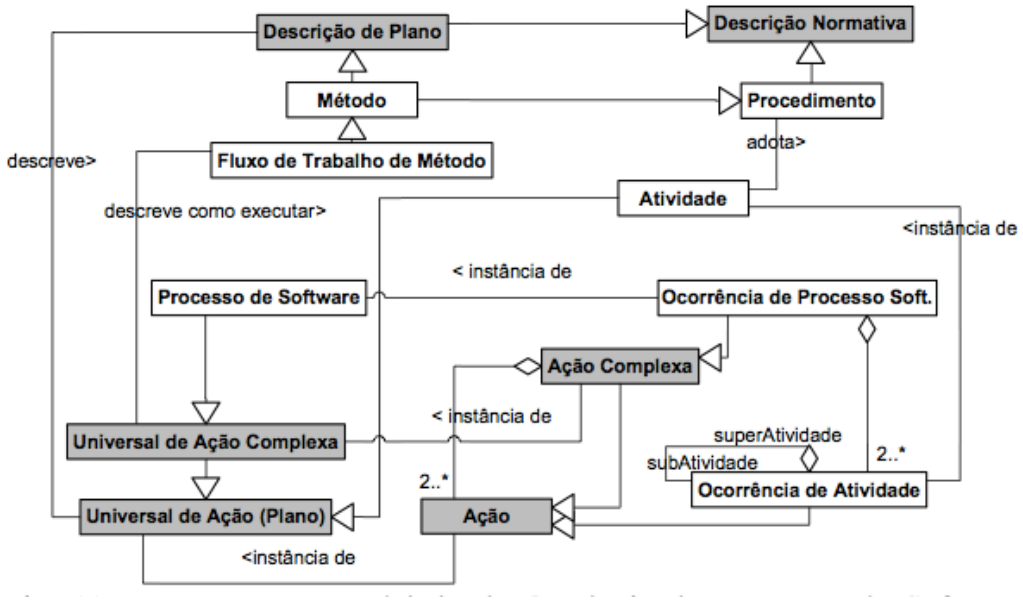

Figura 6 - Fragmento Remodelado da Ontologia de Processo de Software (Eventos). Extraída de [Guizzardi et al, 2008].

O artigo aqui resumido apresentou como UFO pode ser usada para avaliar, re-projetar e dar semântica de mundo real para uma ontologia no domínio de engenharia de software, a saber a ontologia de processo de software que é o cerne do ambiente de desenvolvimento de software *ODE*. Fazendo isso, corrigiu-se um número de problemas conceituais nessa ontologia, tornando-a mais fiel ao domínio representado e tornando seus comprometimentos ontológicos explícitos.

Foi apresentada em [Almeida et al, 2008] uma base semântica para conceitos relacionados a papeis em modelagem empresarial.

Tipicamente, o conceito de papel é usado para definir as responsabilidades e propriedades que se aplicam aos "atores" durante a reprodução de "papéis". E quais ações (ou tipos de ações) são realizadas por que "atores".

"Papeis" são também altamente relevantes quando se discute ações que são realizadas pelos usuários durante a interação com um serviço ou sistema. Neste caso, torna-se necessário definir as ações (tipos de) que podem ser realizados por usuários (tipos de), bem como a representação de usuários (identidades e suas propriedades no âmbito do serviço ou sistema).

A Figura 6 mostra um trecho da ontologia de fundamentação adotada no trabalho.

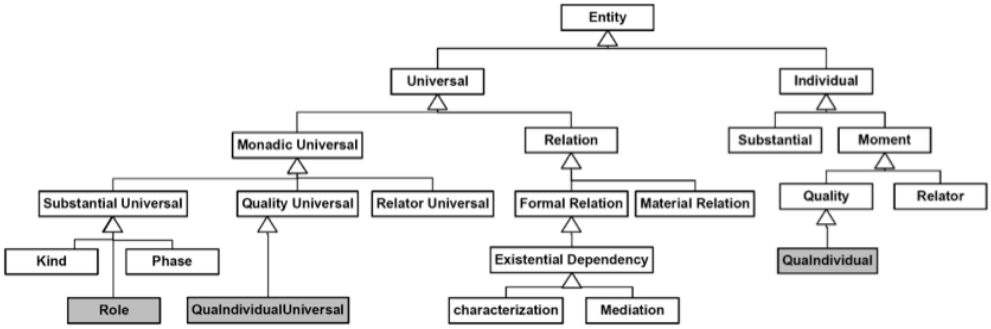

Figura 7 - Extrato da ontologia de fundamentação. Extraída de [Almeida et al, 2008]

O trabalho caracterizou "atores", "agentes" ou "objetos" como substanciais e explicou noções relacionadas a papel. Foram usadas meta-propriedades de universais (ou seja, a dependência existencial, dependência externa e rigidez) para esclarecer certos aspectos dos conceitos relacionados a papel.

2.2 Estado da arte

Esta seção apresenta trabalhos importantes relacionados aos principais temas abordados neste trabalho. É dividida em duas partes: Utilização de dados semânticos em empresas de mídia e Modelos de estruturas narrativas.

2.2.1 Utilização de dados semânticos em empresas de mídia

[Troncy 2008] apresenta uma arquitetura projetada pelo IPTC (*International Press Telecommunication Council*) para facilitar o intercâmbio de notícias.

Como parte dessa arquitetura, são usados vocabulários específicos e controlados, como o *IPTC News Codes*, que visa categorizar itens de notícia (texto, foto, vídeo, etc) de forma padronizada com o vocabulário usado pela indústria.

O objetivo final é criar um ambiente que facilite aos usuários finais verem conexões significativas entre os itens de notícia, seus relacionamentos e os conhecimentos básicos relacionados a eles. Foi criada uma base de conhecimento sobre itens de notícia, além de modelos semânticos de meta-dados para melhorar a interoperabilidade em toda a cadeia de produção de notícias.

Como parte do projeto, foi desenvolvido um guia de como construir uma infra estrutura de Web Semântica para notícias. O primeiro passo é a modelagem da ontologia NAR (*NewsML Architecture*), que é um modelo genérico que define quatro objetos principais (newsItem, packageItem, conceptItem e knowledgeItem). O segundo passo foi interligar essa ontologia com outros padrões da indústria multimídia que já foram convertidos em ontologias OWL (*EXIF*, *Dublin Core*, *XMP*, *DIG35* e *MPEG-7*). O terceiro passo foi converter o *IPTC News Codes* para vocabulário *SKOS* e disponibilizá-lo na web[1] acompanhado de dicas das melhores práticas para publicar vocabulários *RDF*[2] e boas *URIs* para a Web Semântica. O enriquecimento de metadados de notícias foi o quarto e último passo do guia. Foi aplicado processamento linguístico para itens de notícia textuais, e análise visual para itens de notícia de tipos foto e vídeo.

A fim de demonstrar a adequação da infraestrutura da ontologia criada, foi apresentado um ambiente exploratório para a busca e navegação em notícias. Foi utilizado um servidor *web* de busca semântica chamado ClioPatria. O sistema está acessível em http://newsml.cwi.nl/explore/search.

Em [O' Donovan 2010] são descritas as mudanças em tecnologia e fluxo de trabalho usados para gerenciar e publicar o conteúdo da BBC para o site da Copa do Mundo 2010. Este site utiliza as tecnologias da Web Semântica, mais especificamente, "Linked Data" [3], para gerenciar o conteúdo publicado e tem mais de 700 páginas agregadoras de conteúdo (páginas de índice). Por exemplo, a página de um time ou página de um atleta são geradas automaticamente, a partir de "tags" semânticos inseridos nos conteúdos pelos jornalistas. Desta forma, se um conteúdo tem o "tag" "Robinho", é possível, através de inferências, deduzir que este conteúdo é relevante também para a página da seleção brasileira Esta estrutura não seria possível se para cada uma fosse necessário uma intervenção editorial para manter as informações organizadas e atualizadas.

---

[1] http://newsml.cwi.nl/NewsCodes

[2] http://www.w3.org/RDF/ - W3C

[3] http://esw.w3.org/LinkedData

A mudança fundamental é usar métodos avançados para analisar conteúdo e decidir como anotar esse conteúdo com meta-dados precisos, ligados a conceitos que têm identificação única. Um conceito geralmente é uma pessoa, lugar ou artefato do mundo real.

Os princípios por trás desta abordagem são os da fundação da próxima fase da internet, às vezes chamada de Web Semântica, às vezes chamada de Web 3.0. O objetivo é conseguir mais facilmente e com precisão agregar, encontrar e compartilhar conteúdo a partir de várias fontes. Utilizando-se essas relações é possível criar dinamicamente sites e navegações mais ricas em qualquer plataforma.

Há também uma mudança no fluxo de trabalho editorial para a criação de conteúdo e gerenciamento do site. Ao invés de publicar páginas e gerenciar sites o editor publica conteúdo e verifica se as tags sugeridas estão corretas. As páginas de índices são atualizadas automaticamente. Este processo é o que garante a produção de conteúdo de mais alta qualidade e a possibilidade de oferecer com eficiência tantas páginas para a Copa do Mundo.

A figura a seguir descreve de forma geral a arquitetura do site .

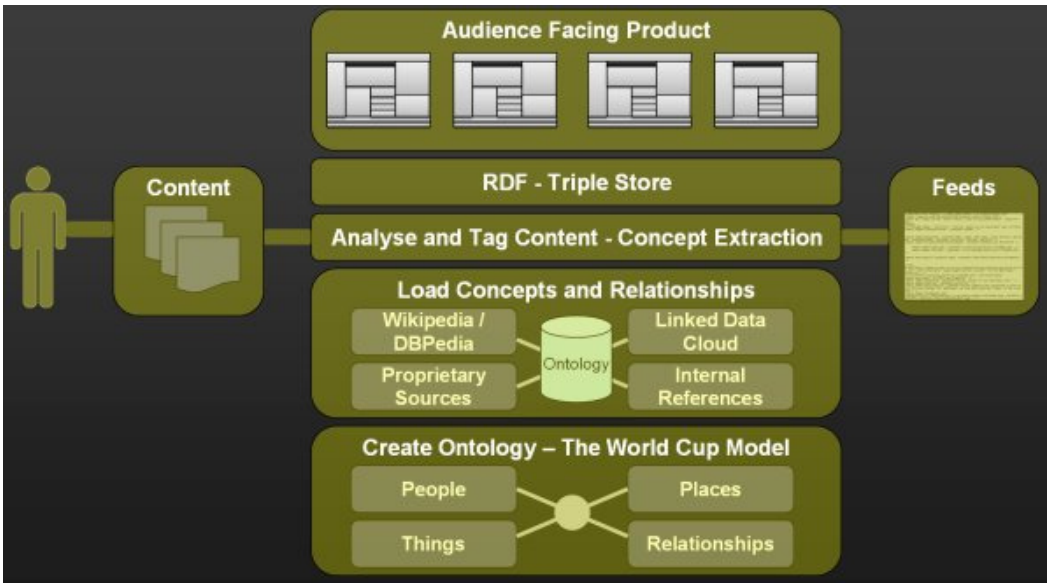

Figura 8 - Arquitetura do site da Copa 2010 da BBC. Extraída de [O' Donovan 2010].

A visão para o futuro na BBC é usar mais tempo na criação e compartilhamento de conteúdo e menos na sua gestão. Para o caso da Copa do Mundo 2010 muitos problemas tiveram de ser superados, principalmente os relacionados com a organização e limpeza de dados. Embora a empresa ainda não disponibilize todo o seu conteúdo em *RDF* ela começará a fazê-lo em breve.

[Dunn 2010] apresenta como o Guardian utilizou conteúdo, funcionalidades de busca e software *open source* para construir um modelo de negócio novo e poderoso.

O projeto chamado Open Platform é dividido em quatro módulos. Content API, Data Store, Politics API e MicroApps. Content API é um serviço para selecionar e coletar conteúdos do Guardian para reutilização. Data Store é um diretório de dados úteis tratados por editores do Guardian. Politics API é um banco de dados de candidatos, registros eleitorais, resultados de eleições e dados em tempo real no dia da eleição. MicroApps é um framework para integrar aplicações de terceiros no guardian.co.uk.

Content API é um exemplo de aplicação que não utiliza banco de dados relacional. Toda aplicação foi feita através de consultas ao índice do Solr (mecanismo

de busca)[4]. De acordo com o artigo publicado foi possível aplicar consultas multi-facetadas mais rápidas em uma arquitetura de busca do que num SGBDR.

Foram definidos níveis de acesso e modelos de rentabilidade entre o Guardian e seus parceiros. Seguindo uma estratégia aberta, que visa trazer para dentro da empresa dados e aplicações da Internet e também permitir que parceiros construam aplicações para outras plataformas digitais usando dados e serviços do Guardian.

[Ó Crualaoich 2010] descreve alguns novos recursos de Linked Data do Guardian OpenPlatform.

O conteúdo do Guardian Content API foi estendido para incluir identificadores de outras bases de dados externas. No momento, existem dois tipos de identificadores externos com os quais os dados do Guardian estão relacionados: ids do ISBN e do MusicBrainz. Os ids do ISBN estão disponíveis em aproximadamente 2700 críticas literárias. Existe um total de aproximadamente 17000 artigos com potencial de uso dos ids ISBN e já está sendo feito um trabalho nesse sentido. Para as novas críticas literárias, todas devem conter o ISBN. Os ids MusicBrainz estão disponíveis em cerca de 17000 itens de conteúdo. Devido ao modelo de domínio do Guardian tratar artistas e bandas como objetos primários, com sua própria tag associada, é muito mais fácil para anotar ids MusicBrainz. Existem cerca de 600 artistas que foram anotados dessa maneira.

Em [New York Times 2009] é descrito o lançamento de 5000 tags na *Linked Open Data (LOD)*[5].

Durante meses foram mapeadas manualmente 5000 tags de pessoas da base do New York Times com as bases da DBPedia e da Freebase. Agora é possível acessar a URI de cada pessoa (segundo a base do NYT) e verificar a sua instância equivalente nessas outras duas bases. É possível também acessar a página de tópico de uma pessoa no site nytimes.com. O mais importante é que computadores podem acessar a URI de uma pessoa e obter todas essas informações em *RDF*.

---

[4] http://lucene.apache.org/solr/

[5] http://esw.w3.org/LinkedData - W3C

Antes dessa integração a base de assuntos era muito isolada. Mesmo que fosse possível exibir a lista de artigos sobre uma determinada pessoa, por exemplo, não era possível exibir a data de nascimento dessa pessoa pois esta informação não existia na base de tags.

Para garantir que os dados serão usados da forma mais ampla e livre possível o NYT anunciou que todos os registros de dados disponíveis em *data.nytimes.com* serão publicados sob a licença Creative Commons 3.0 Attribution License.

2.2.2 Modelos de estruturas narrativas

A pesquisa relatada em [Ciarlini et al, 2009] descreve a composição de narrativas no contexto de contar histórias interativamente. Um dos pontos centrais dessa pesquisa é a relação de eventos na composição de enredo baseada em planos. Foi desenvolvido um novo tipo de narrativa para jogos: não-linear, orientada pelo jogo, que gira em torno da experiência do jogador, com o objetivo de fazer o jogador deixar de ser um mero consumidor da narrativa para ser consumidor e coautor.

Para apoiar a produção de histórias, é utilizado o que a pesquisa semiótica destacou como os quatro principais "tropos": metáfora, metonímia, sinédoque e ironia. Foram mapeados e analisados alguns tipos de relação entre eventos de uma história: relações paradigmáticas , relações metonímicas, relações sintagmáticas e relações antitéticas. Pesquisas de narratologia distinguem três níveis de composição literária, fabula, história e texto. O trabalho foi realizado para narrativas no nível de fabula e considerou sete papeis distintos para personagens, de acordo com os acontecimentos atribuídos à iniciativa de cada um. São eles: o herói, o vilão, a vítima, o despachante, o doador, o ajudante e o falso herói.

Alguns sistemas e abordagens computacionais foram usados para apoiar a narrativa interativa. Algoritmos de Planejamento provaram ser uma alternativa útil para ajudar a criar narrativas, explorando diferentes cadeias de eventos para a realização dos objetivos dos personagens. Foram projetados três esquemas contextuais: esquema estático, que especifica as entidades, relacionamentos e atributos; esquema dinâmico, que define um repertório fixo de operações para executar de forma consistente as mudanças de estado e esquema comportamental, que consiste de regras de inferência de objetivo, regras de crenças e as regras de condição

emocional. Um protótipo simples, PlotBoard, foi projetado para experimentar os conceitos discutidos.

A interação das relações paradigmáticas, metonímicas, sintagmáticas e antitéticas já permitiu uma ampla cobertura que foi reforçada pela ligação entre essas relações e os quatro tropos principais. Somente o nível de fábula foi abordado onde foi possível indicar quais eventos devem ser incluídos na narrativa. Um problema complexo a ser enfrentado será o nível da história, onde a preocupação é como contar os eventos.

Nota-se que o trabalho sobre composição de narrativas no contexto de games utiliza modelos de estruturas narrativas, porém, os trabalhos realizados no contexto de mídia jornalística ainda não o utilizam.