



**Rafael Antônio Pinto Pena**

**Suporte semântico à publicação de  
conteúdo jornalístico na Web**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio.

Orientador: Prof. Daniel Schwabe

Rio de Janeiro  
Fevereiro de 2012



**Rafael Antônio Pinto Pena**

**Suporte semântico à publicação de  
conteúdo jornalístico na Web**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Daniel Schwabe**

Orientador e Presidente

Departamento de informática - PUC-Rio

**Prof. Edward Hermann Haeusler**

Departamento de informática - PUC-Rio

**Prof. Giancarlo Guizzardi**

Departamento de informática - UFES

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 07 de Fevereiro de 2012

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

### **Rafael Antônio Pinto Pena**

Graduou-se em Tecnólogo em Processamento e Dados pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ) em Dezembro de 1999. Tem experiência na área de Ciência da Computação, com ênfase em Desenvolvimento de Software. Tem trabalhado em análise de sistemas desde 1997. Mais especificamente com Desenvolvimento de Sistemas para Web, e recentemente, Desenvolvimento de Produtos Jornalísticos para Web.

#### Ficha Catalográfica

Pena, Rafael Antônio Pinto

Suporte semântico à publicação de conteúdo jornalístico na Web / Rafael Antônio Pinto Pena; Orientador: Daniel Schwabe. - Rio de Janeiro: PUC-Rio, Departamento de Informática, 2012.

v., 105 f.: il. ; 29,7 cm

Dissertação de Mestrado - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática

Referencias bibliográficas incluídas.

Web Semântica; Modelo de Domínio; Conteúdo Jornalístico; Conteúdo Web; Empresas Jornalísticas; Mídia Digital

## **Agradecimentos**

A Deus pela maravilhosa forma de vida que me foi concedida.

A minha esposa Carolyn Maciel pelo apoio, amor, compreensão e por estar ao meu lado em todos os momentos me tranquilizando e fazendo acreditar que seria possível.

Aos meus pais, Fernando Pena e Nina Pena, pela educação, carinho, amor e por me mostrarem a importância do aprendizado e do trabalho.

As minhas irmãs Catarina Pena e Fabiana Pena Ruiz pelo amor, pelos exemplos e ensinamentos.

A minha avó Griselda Schleder, por lutar contra a doença para permanecer viva até ver seus netos formados. Uma inesquecível demonstração de amor e de valorização do estudo.

Ao meu orientador Professor Daniel Schwabe, pelo enorme apoio, confiança e orientação.

Ao Professor Giancarlo Guizzardi pela grande ajuda na modelagem conceitual.

Ao amigo Janilton Maciel pela enorme ajuda no desenvolvimento do protótipo.

Aos amigos da Globo.com, Davidson Fellipe, Franklin Amorim, Claudio Braga, Paulo Ricardo Muniz, Camila Dias, Newton Fleury e Alexandre Moreno por sempre me apoiarem e colaborarem com ideias e melhorias.

Sou grato aos jornalistas que participaram da pesquisa e tanto ajudaram neste trabalho. André Amaral, Gustavo Poli, Bernardo Ferreira, Bernardo Pombo, Eduardo Orgler e Rodrigo Breves.

À Globo.com, pelo financiamento e auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado. Em especial ao Juarez Queiroz por conseguir este benefício para os funcionários e ao Jacques Douglas Varaschim por todo apoio.

Aos professores da Comissão examinadora.

## Resumo

Pena, Rafael Antônio Pinto; Schwabe, Daniel. **Suporte semântico à publicação de conteúdo jornalístico na Web**. Rio de Janeiro, 2012. 105p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O consumo de conteúdo jornalístico na Web aumenta a cada dia. No entanto, boa parte desse conteúdo ainda é produzido segundo paradigmas da mídia impressa. Paralelamente a isto, a Web Semântica ou Web 3.0, adiciona uma camada de "inteligência" à Web, onde computadores são capazes de extrair significados dos conteúdos acessados na Web, e consequentemente, processá-los. Neste trabalho foi definido e testado um modelo de publicação de conteúdo jornalístico apoiado pela Web Semântica. Foi desenvolvida uma ferramenta de apoio para produtores de conteúdo, com o objetivo de melhorar o processo de construção de narrativas jornalísticas para web, através da sugestão de padrões narrativos e de informações objetivas que dão suporte ao tipo de narrativa desejado. Um estudo de caso foi realizado para avaliar em ambiente real o modelo proposto. Um grupo de jornalistas usou a ferramenta desenvolvida, com aceitação unânime.

## Palavras-chave

Web Semântica; Modelo de Domínio; Conteúdo Jornalístico; Conteúdo Web; Empresas Jornalísticas; Mídia Digital.

## **Abstract**

Pena, Rafael Antônio Pinto; Schwabe, Daniel (Advisor) . **Semantic Support for the publication of news content on the Web**. Rio de Janeiro, 2011. 105p. MSc Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The use of journalistic content on the web increases every day. However, much of the content is still produced according to the paradigms of print media. Parallel to this, the Semantic Web or Web 3.0, adds a layer of "intelligence" to the Web, where computers are able to extract meaning from the visited web content, and consequently, process them. This work defined and tested a model of publishing news content supported by the Semantic Web. It was developed a support tool for content producers, in order to improve the process of construction of news stories for the web, using the suggestion of narrative patterns and objective information that support the desired type of narrative. A case study was conducted to evaluate the proposed model in a real environment. A group of journalists used the developed tool, with unanimous acceptance.

## **Keywords**

Semantic Web; Domain Model; Journalistic Content; Web Content; Media Companies; Digital Media.

# Sumário

1 Introdução	12
2 Fundamentos e estado da arte	14
2.1 Fundamentos	14
2.1.1 Web Semântica e Linked Open Data	14
2.1.2 Sistema de frames e slots	20
2.1.3 UFO ( <i>Unified Foundational Ontology</i> )	22
2.2 Estado da arte	32
2.2.1 Utilização de dados semânticos em empresas de mídia	32
2.2.2 Modelos de estruturas narrativas	37
3 Modelo de publicação de conteúdo jornalístico na web	39
3.1 Descrição do contexto de uso	39
3.2 Descrição do fluxo desejado	41
4 Modelos de ontologias	45
4.1 Descrição da metodologia de modelagem usada	45
4.1.1 Construção de um modelo de jogo de futebol	46
4.1.2 Definição dos estereótipos e esquemas narrativos	48
4.2 Ontologia do Futebol	51
4.2.1 Fragmento das participações em um jogo	52
4.2.2 Fragmento das etapas de disputa de um jogo	54
4.2.3 Fragmento do lance de disputa comum	55
4.2.4 Fragmento do lance de jogo individual	58
4.2.5 Fragmento dos Lances Complexos	61
4.2.5.1 Lances faltosos	61
4.2.5.2 Lances de jogo	63
4.2.6 Resumo do modelo	66
4.3 Modelo de estereótipos e narrativas	68
4.3.1 Regras de detecção de estereótipos	68

4.3.2 Formação dos esquemas narrativos	75
5 Estudo de caso	80
5.1 Protótipo funcional	80
5.1.1 Interface de autoria de matéria	80
5.1.2 Exemplos de templates de estereótipos	85
5.2 A experiência de uso	91
5.2.1 Indicadores sobre a utilização do protótipo	92
5.2.2 Pesquisa de opinião sobre o uso do protótipo	96
6 Conclusões e próximos passos	98
6.1 Conclusões	98
6.2 Próximos passos	100
6.2.1 Oportunidades relacionadas ao modelo de publicação	100
6.2.2 Oportunidades relacionadas aos esquemas narrativos	101
7 Referências bibliográficas	103



## Lista de imagens

Figura 1 – Quadrado de Aristóteles. Figura extraída de [Guizzardi et al, 2009]	23
Figura 2 - Fragmento da UFO-B. Objetos e Eventos. Figura extraída de [Guizzardi et al, 2008]	26
Figura 3 - Fragmento da UFO-B: Relações de Allen. Figura extraída de [Guizzardi et al, 2008]	27
Figura 4 - Fragmento da Ontologia de Processo de Software de ODE. Extraída de [Guizzardi et al, 2008]	29
Figura 5 - Fragmento Remodelado da Ontologia de Processo de Software (Substanciais). Extraída de [Guizzardi et al, 2008]	29
Figura 6 - Fragmento Remodelado da Ontologia de Processo de Software (Eventos). Extraída de [Guizzardi et al, 2008]	31
Figura 7 - Extrato da ontologia de fundamentação. Extraída de [Almeida et al, 2008]	32
Figura 9 - Esquema simplificado do processo de publicação de conteúdo web existente	39
Figura 10 - Esquema simplificado do processo de publicação de conteúdo web proposto no trabalho	43
Figura 11 - Fragmento das participações em um jogo	52
Figura 12 - Fragmento das participações em um jogo alinhado com a UFO	53
Figura 13 - Fragmento das etapas de disputa de um jogo	54
Figura 14 - Fragmento das etapas de disputa de um jogo alinhado com a UFO	55
Figura 15 - Fragmento do lance de disputa comum	56
Figura 16 - Fragmento do lance de disputa comum alinhado com a UFO	57

Figura 17 - Fragmento do lance de jogo individual	58
Figura 18 - Fragmento do lance de jogo individual alinhado com a UFO	61
Figura 19 – Fragmento do lances faltosos	62
Figura 20 – Fragmento dos lances faltosos alinhado com a UFO	63
Figura 21 – Fragmento do lances válidos	64
Figura 22 – Fragmento dos lances válidos alinhado com a UFO	66
Figura 23 - Matriz de estereótipos	68
Figura 24 - Interface de autoria de matérias	82
Figura 25 - Interface para seleção de jogo para crônica	83
Figura 26 - Interface de sugestão de estereótipos	84
Figura 27 - Interface de autoria de matéria com template de estereótipo selecionado	85
Figura 28 - Template narrativo para estereótipo de jogo movimentado	86
Figura 29 - Template narrativo para estereótipo de jogo morno	86
Figura 30 - Template narrativo para estereótipo de goleada	87
Figura 31 - Template narrativo para estereótipo de placar duvidoso	88
Figura 32 - Trecho de matéria feita com template Goleada	88
Figura 33 - Trecho de matéria feita com template Placar Duvidoso	89
Figura 34 - Trecho de matéria feita com template Artilheiro	90
Figura 35 - Gráfico comparativo entre quantidade de matérias produzidas e quantidade de matérias produzidas com templates narrativos do protótipo	92
Figura 36 - Gráfico comparativo entre quantidade de jogos e a quantidade de jogos para os quais pelo menos uma matéria foi produzida com template narrativo do protótipo	94
Figura 37 – Gráfico com as quantidades de uso dos templates narrativos do protótipo	95

## Lista de Tabelas

Tabela 1 - Estereótipo Jogo Movimentado	69
Tabela 2 - Estereótipo Jogo Morno	70
Tabela 3 - Estereótipo Jogo Truncado	71
Tabela 4 - Estereótipo Goleada	71
Tabela 5 - Estereótipo Jogo Recorde no campeonato	72
Tabela 6 - Estereótipo Placar Duvidoso	73
Tabela 7 - Estereótipo Melhor Performande do Jogo	75
Tabela 8 - Template narrativo Jogo Movimentado / Jogo Morno	75
Tabela 9 - Template narrativo Goleada	76
Tabela 10 - Template narrativo Jogo Truncado	77
Tabela 11 - Template narrativo Placar Duvidoso	77
Tabela 12 - Template narrativo Primeiro no ranking do campeonato	78
Tabela 13 - Template narrativo Atleta em Destaque	79
Tabela 14 - Resultado da pesquisa de opinião feitas com os jornalistas envolvidos	97

## 1 Introdução

Há mais de 40 anos, Herbert Marshall McLuhan, professor e filósofo canadense, deu origem à expressão "o meio é a mensagem" [McLuhan 1964]. No seu argumento, novas tecnologias (como alfabetos, prensas de impressão e até a fala em si) exercem um efeito sobre a cognição, que por sua vez influencia a organização social e a forma como as pessoas se comunicam.

Segundo McLuhan, "mensagem" é "a mudança de escala, ritmo ou padrão" que uma nova invenção ou inovação "introduz nas questões humanas". A Internet, mais especificamente a WWW, são claramente "mensagens" neste sentido e se constituem novas mídias. Por extensão, a Web Semântica [Berners-Lee et al, 2001] ou Web 3.0 sendo um novo formato de conteúdo para Web, onde computadores são capazes de extrair significados dos conteúdos acessados na Web, e consequentemente, processá-los, é também uma nova mídia.

Diante desse cenário, as empresas jornalísticas tem pela frente o desafio de rever a forma como produzem conteúdo para Web, a fim de adaptar a mensagem a este novo meio. Nesta iniciativa, surge a oportunidade de se utilizar os recursos da Web Semântica para melhorar a cadeia produtiva e entregar para o usuário final uma experiência mais rica.

Este trabalho de pesquisa teve como principal objetivo definir e testar um modelo de publicação de conteúdo jornalístico apoiado pela Web Semântica. O modelo serviu como ferramenta de apoio para produtores de conteúdo durante o processo narrativo, auxiliando na melhoria do processo de produção e na criação de conteúdos mais ricos em significado e diversidade. Mais especificamente, foram utilizadas representações de modelos de narrativa que permitiram a composição de forma semiautomática de matérias jornalísticas em mídia digital.

O modelo tem como base as características do domínio do conteúdo a ser produzido e as características das estruturas narrativas adequadas para este domínio.

A solução foi especializada, inicialmente, para o domínio de futebol. Mais especificamente, para dar suporte a criação de narrativas acerca de um jogo. Uma ontologia de futebol foi criada para representar de forma semanticamente correta um jogo de futebol. Desta forma, foi possível mapear os eventos, agentes e participações possíveis em um jogo. Para experimentar o processo de autoria de conteúdo web sugerido neste trabalho foi desenvolvida uma ferramenta que identifica personagens, fatos importantes e outros fatos do domínio de futebol relacionados aos acontecimentos do jogo e disponibiliza essas informações para o produtor de conteúdo durante o processo de criação de matérias.

No capítulo 2 serão apresentados os principais fundamentos usados neste trabalho e os principais trabalhos relacionados que foram identificados. Conceitos sobre Web Semântica, *Linked Data*, Ontologias e Modelagem de Ontologias serão apresentados neste capítulo.

O terceiro capítulo, Modelo de publicação de conteúdo jornalístico na web, apresenta de forma simplificada, o modelo de publicação usado na organização onde a pesquisa foi realizada e o novo modelo proposto neste trabalho.

No quarto capítulo, Modelos de ontologias, é apresentada a ontologia de futebol e os modelos de estereótipos e esquemas narrativos desenvolvidos, assim como as metodologias usadas na construção desses artefatos.

O capítulo 5 descreve o protótipo desenvolvido e como foi o experimento realizado por um grupo de jornalistas em ambiente real.

No sexto capítulo, Conclusões e próximos passos, são apresentadas as principais conclusões sobre o trabalho e as oportunidades de trabalhos futuros relacionados.

Por último, no capítulo 7 Referências bibliográficas, são apresentadas as referências bibliográficas usadas.

## 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 apresentada 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.

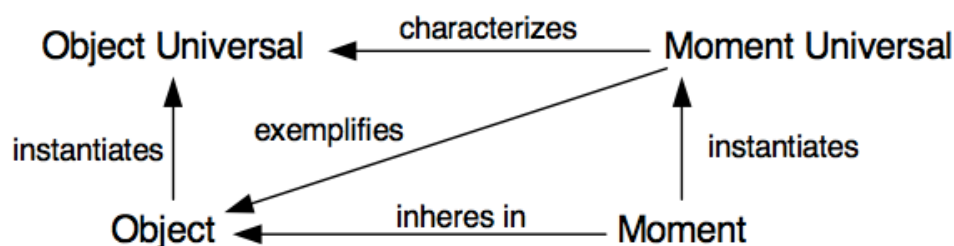


Figura 1 – Quadrado de Aristóteles. Figura extraída de [Guizzardi et al, 2009].

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 *Momento* 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).

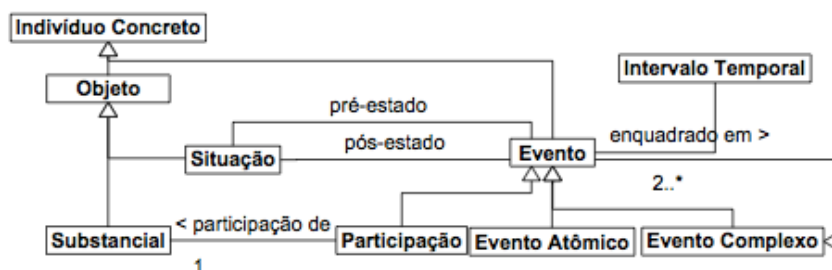


Figura 2 - Fragmento da UFO-B. Objetos e Eventos. Figura extraída de [Guizzardi et al, 2008]

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.

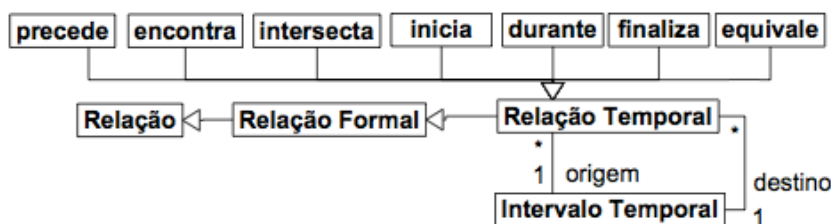


Figura 3 - Fragmento da UFO-B: Relações de Allen. Figura extraída de [Guizzardi et al, 2008].

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 comprometicimentos 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.

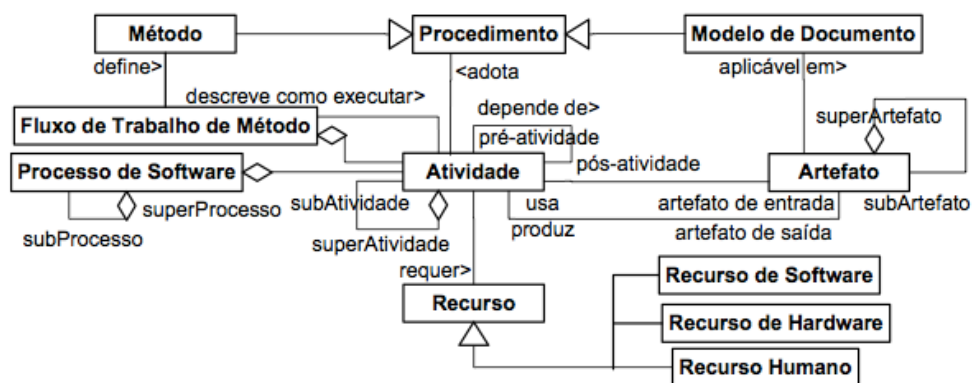


Figura 4 - Fragmento da Ontologia de Processo de Software de ODE. Extraída de [Guizzardi et al, 2008].

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.

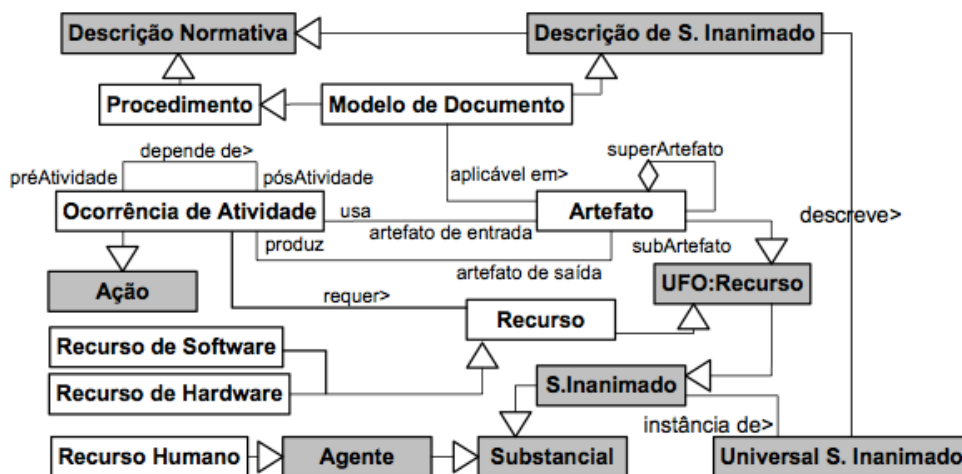


Figura 5 - Fragmento Remodelado da Ontologia de Processo de Software (Substanciais). Extraída de [Guizzardi et al, 2008].

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.

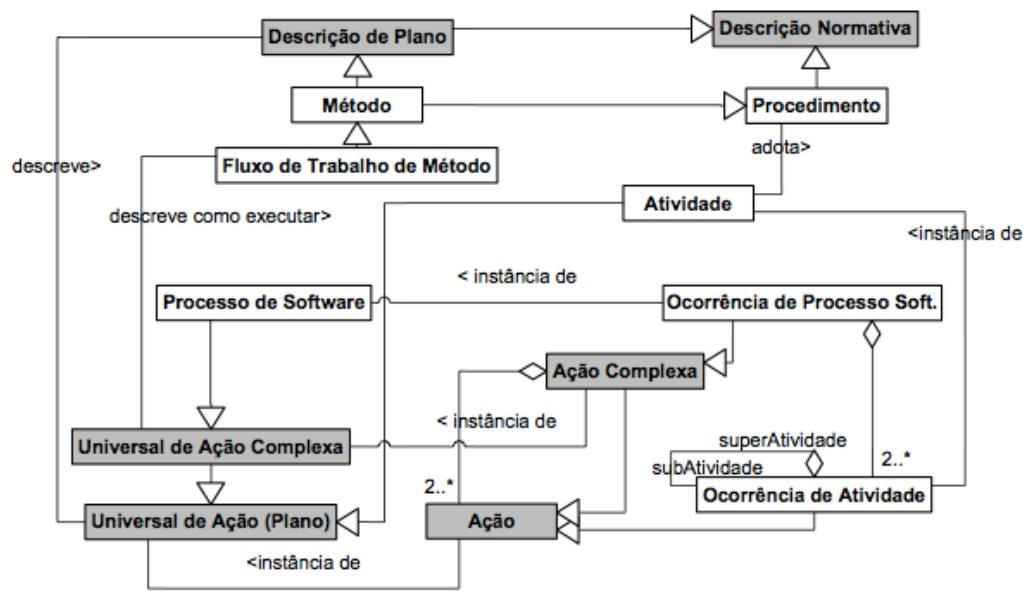


Figura 6 - Fragmento Remodelado da Ontologia de Processo de Software (Eventos). Extraída de [Guizzardi et al, 2008].

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 papéis 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".

"Papéis" 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.

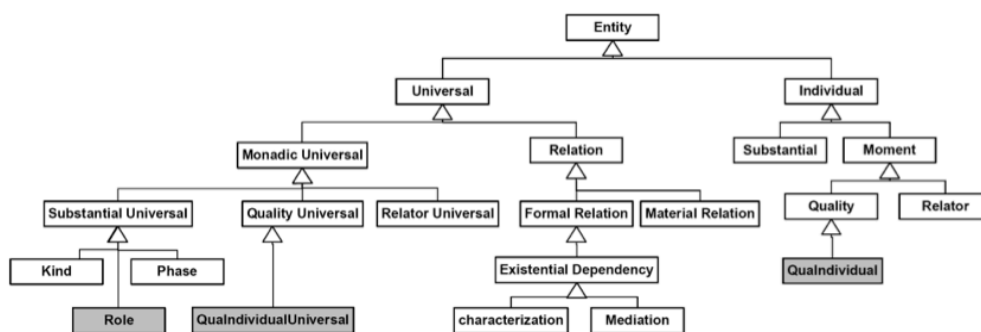


Figura 7 - Extrato da ontologia de fundamentação. Extraída de [Almeida et al, 2008]

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 infraestrutura 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<sup>1</sup> acompanhado de dicas das melhores práticas para publicar vocabulários *RDF*<sup>2</sup> 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”<sup>3</sup>, 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.

---

<sup>1</sup> <http://newsml.cwi.nl/NewsCodes>

<sup>2</sup> <http://www.w3.org/RDF/> - W3C

<sup>3</sup> <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 .

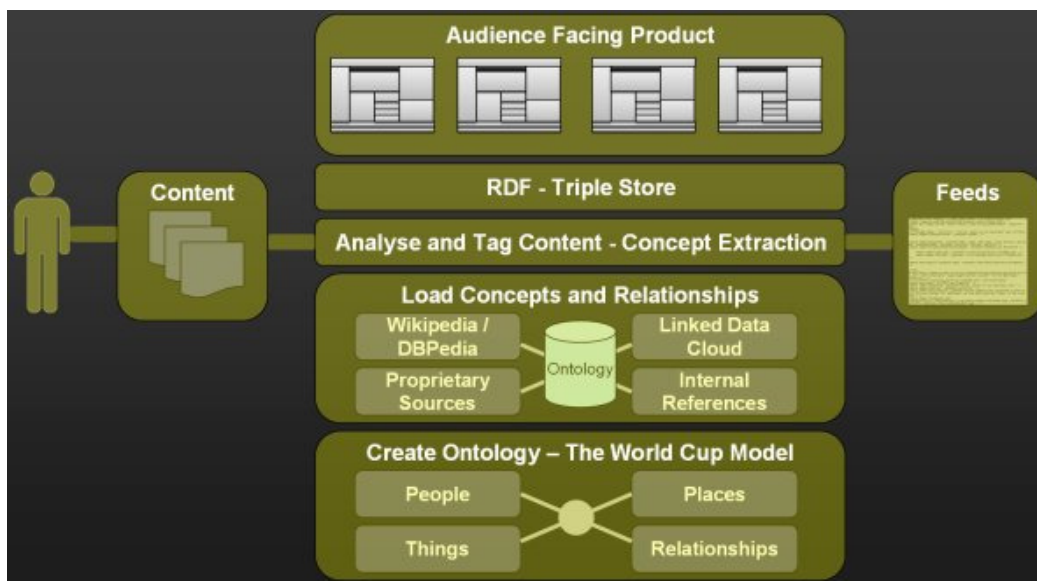


Figura 8 - Arquitetura do site da Copa 2010 da BBC. Extraída de [O' Donovan 2010].

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)<sup>4</sup>. De acordo com o artigo publicado foi possível aplicar consultas multifacetadas 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 CruaIaoich 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)*<sup>5</sup>.

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*.

---

<sup>4</sup> <http://lucene.apache.org/solr/>

<sup>5</sup> <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 papéis 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.

### 3 Modelo de publicação de conteúdo jornalístico na web

Este capítulo tem como objetivo mostrar quais as diferenças entre o modelo de publicação de conteúdo para *web* proposto neste trabalho e o modelo existente na organização onde a pesquisa foi realizada.

#### 3.1 Descrição do contexto de uso

Esta seção é dedicada à descrição de como é o processo de produção de matérias e conteúdos para web, atualmente, dentro da organização onde o experimento foi realizado. Aqui serão explicados os principais papéis envolvidos nesse processo, quais as suas contribuições dentro da cadeia de produção e como eles colaboram entre si. A Figura 9 apresenta um esquema simplificado do processo de publicação de conteúdo e seus agentes.

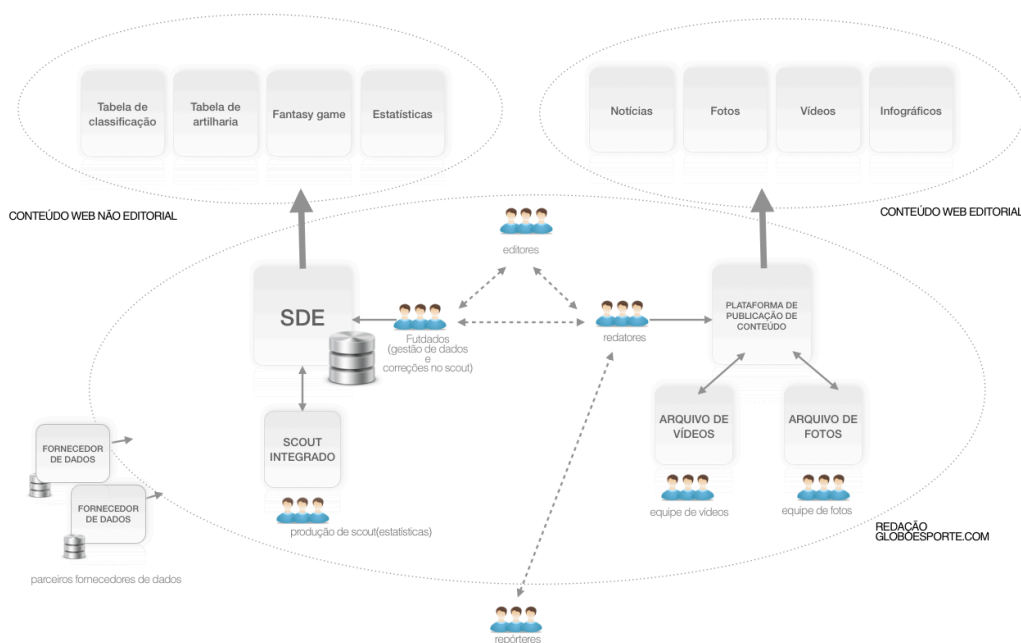


Figura 9 - Esquema simplificado do processo de publicação de conteúdo web existente.

Atualmente, é comum encontrar dentro de uma organização de mídia, que produz conteúdo para web, uma nova função, o editor de dados, exercida por profissionais responsáveis por manter bases de dados atualizadas ao longo do tempo.

No ambiente analisado, uma organização que produz conteúdo esportivo, esse profissional faz uso de sistemas para realizar tarefas como, por exemplo, cadastrar campeonatos e jogos, manter os elencos dos times atualizados, cadastrar resultados de jogos, conferir e importar resultados e estatísticas (faltas, gols, defesas, etc) de jogos fornecidos por parceiros. A equipe responsável por manter atualizada a base de dados de futebol se chama Futdados. O sistema utilizado pela equipe é chamado SDE (Sistema de Dados Esportivos).

Algumas informações disponibilizadas na *web* são oriundas dessas bases de dados. Tabelas de classificação, tabelas de artilharia e seção de estatísticas são alguns exemplos. As páginas *web* onde esses conteúdos são expostos para os internautas são atualizadas, automaticamente, sempre que as bases de dados são alteradas.

O editor estabelece uma linha editorial a ser seguida pelos redatores e repórteres. É o responsável pelo direcionamento dos assuntos sobre os quais os redatores produzirão o conteúdo. Ele direciona o destaque que é dado para o conteúdo estabelecendo critérios de importância dos assuntos. Embora também produza conteúdo, sua principal função é direcionar e acompanhar a produção de conteúdo por parte dos redatores e repórteres.

Apurar a informação, seja oriunda de um repórter, ou de um veículo de informação qualquer é responsabilidade do redator. Ele apura a informação e produz o conteúdo que será disponibilizado para os internautas. É o usuário mais frequente do sistema de publicação de conteúdo. Com esse sistema o redator consegue criar matérias utilizando textos, infográficos, fotos e vídeos. Quando o redator quer escrever alguma matéria de pesquisa, que fale sobre o jogo com o maior número de faltas do campeonato, por exemplo, ele entra em contato com a equipe Futdados para obter informações sobre estatísticas de faltas nos jogos do campeonato. Com essas informações e mais os recursos disponíveis no sistema de publicação ele produz o conteúdo final.

O repórter é quem atua fora da redação buscando novas informações. Ele trabalha, remotamente, enviando para o redator informações sobre clubes, atletas, competições, enfim, pessoas, eventos e organizações relacionadas ao esporte. Muitas vezes essas informações são obtidas nos estádios ou nos clubes. As vezes



o repórter usa o sistema de publicação para cadastrar conteúdo e em seguida avisa ao redator que o revisa. Esse fluxo acontece sempre que o repórter tem pressa para dar um furo. Portanto, o repórter também é usuário do sistema de publicação. Porém, é mais frequente o repórter enviar o conteúdo para o redator que apura e depois publica para o público na Internet.

Existe também uma equipe responsável pelas fotos usadas nos conteúdos. Esses profissionais trabalham dando suporte aos redatores, inserindo no sistema de publicação as fotos que os mesmos precisam. Sempre que algum redator precisar de uma foto específica essa equipe é acionada. A equipe verifica se já existe alguma foto no sistema que obedeça aos critérios informados pelo redator e caso exista, informa para o mesmo a descrição da foto para que ele consiga localizá-la no sistema. Caso ainda não exista, a equipe de fotos cadastra uma nova foto no sistema para em seguida avisar ao redator. Essa equipe é responsável por organizar as fotos fornecidas pelas agências contratadas. Em geral, essas agências disponibilizam um sistema web, através do qual o cliente pode baixar as fotos que têm direito mediante autenticação no sistema.

Podemos encontrar algumas outras funções dentro da redação que não participam, ativamente, do processo de criação de matérias e conteúdos para web e por isso não foram descritas nessa seção.

### **3.2 Descrição do fluxo desejado**

O fluxo proposto neste trabalho é baseado em um modelo que tem dois objetivos principais: melhorar o processo de produção e viabilizar um novo tipo de oferta de conteúdos, de forma semiautomática, alavancado por modelos semânticos com suporte das tecnologias da Web Semântica. Este modelo terá como base as características do domínio do conteúdo a ser produzido e as características das estruturas narrativas adequadas para este domínio.

Do ponto de vista prático, o modelo servirá para melhorar o processo de construção de narrativas jornalísticas para web, através da sugestão de padrões narrativos e de informações objetivas que dão suporte ao tipo de narrativa desejado. A solução será especializada, inicialmente, para o domínio de futebol. O protótipo

desenvolvido tem como objetivo dar suporte à criação de matérias contendo narrativas ligadas a eventos de futebol.

Ao criar uma matéria acerca de um jogo o sistema deverá identificar personagens, fatos importantes e outros fatos do domínio de futebol relacionados aos acontecimentos do jogo. É importante ressaltar que a ferramenta proposta não é um simples agregador de conteúdo. A ferramenta deve entender padrões narrativos, identificar esses padrões ao analisar informações do jogo, e agregar informações relevantes de acordo com os padrões identificados. Por exemplo, se um jogo teve muitas faltas, no momento da criação de uma matéria sobre o jogo, o sistema deve sugerir um *template* de “jogo truncado” que exibirá a quantidade de faltas e cartões ocorridos no jogo, assim como, as médias de faltas e cartões por partida no campeonato. Se algum personagem for identificado, por exemplo, um garçom (jogador que dá assistência), o sistema deverá sugerir um *template* de garçom que trará informações como a quantidade de assistências dele no jogo e a sua média de assistências no campeonato.

O sistema de publicação de conteúdo proposto contará com uma duas bases de conhecimento: base semântica de futebol e base semântica de esquemas narrativos. Além disso, as bases de conhecimento utilizadas pelo sistema devem estar conectadas com a LOD para que as matérias sejam enriquecidas também com dados externos ao sistema. Alguns exemplos de dados que podem ser aproveitados são as informações dos estádios de futebol segundo a Wikipedia , e também informações geográficas segundo a *Geonames* .

A Figura 10 descreve de forma geral o fluxo de publicação de conteúdo proposto pelo modelo, e respectivos agentes. Lembrando que esse fluxo está relacionado à publicação de conteúdo editorial. Ou seja, a parte de conteúdo web não editorial, mostrada na seção anterior não será afetada pelo modelo proposto.

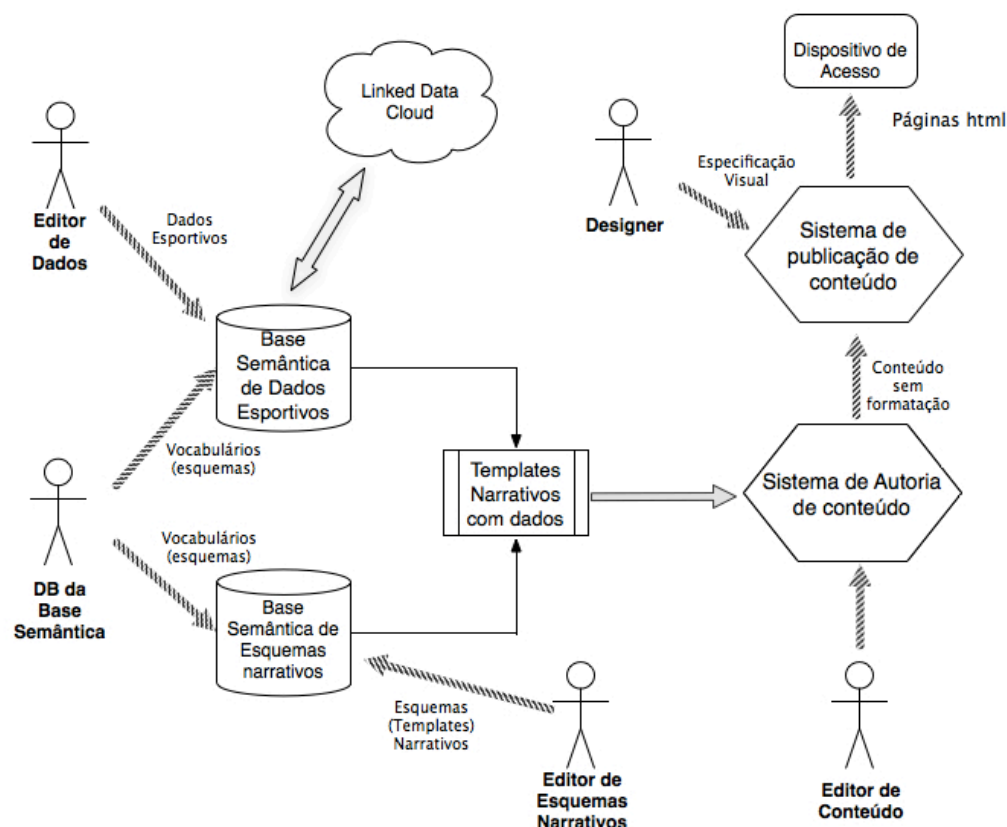


Figura 10 - Esquema simplificado do processo de publicação de conteúdo web proposto no trabalho.

O editor de dados é uma função que já existe (equipe Futdados). Ele é responsável por manter a base de dados esportivos. Esse profissional, em geral, não cria matérias.

Editor de esquemas narrativos será responsável por definir os padrões narrativos que devem ser usados. Deve ser um especialista tanto em futebol quanto em jornalismo. Como resultado do trabalho dele é esperado uma definição dos padrões narrativos nos quais o sistema deve se basear. Ele que define, por exemplo, que o *template* de jogo “truncado” deve apresentar a quantidade de faltas e cartões ocorridos no jogo, junto com as médias de faltas e cartões do campeonato para que o internauta possa fazer uma comparação e entender melhor a informação.

O editor de conteúdo (ou redator) passará a receber sugestões de narrativas (templates narrativos com dados sobre futebol) que o auxiliarão durante o processo de autoria. O sistema ajudará a padronizar narrativas de jogo e a coletar dados relevantes para as mesmas. A ideia é o editor do conteúdo gastar a maior parte do

tempo usando a linguagem que domina para trazer mais emoção ao texto. Enfim, fazer o que não é possível ser feito pela máquina.

Os diferentes agentes da figura atuam em fluxos de trabalho distintos, com diferentes ciclos de utilização do sistema. O designer interage com o sistema quando alguma modificação visual na estrutura da página for necessária. Em geral, define-se uma apresentação e depois todas as matérias publicadas seguem esta apresentação. O editor de esquemas narrativos e o DBA da base semântica também não interagem com o sistema a cada conteúdo produzido. Somente quando for necessário alterar o modelo usado pela ferramenta é que estes agentes utilizam o sistema. Por exemplo, quando é identificado que um determinado dado deve passar a ser registrado acerca de uma partida, para coleta de estatísticas, o DBA atualiza o modelo de dados da base semântica. De forma análoga, o editor de esquemas narrativos atualiza a base quando um novo esquema é identificado (ou criado), assim como quando há necessidade de alterações em algum esquema. Já o editor de dados esportivos e o jornalista que cria as matérias são os agentes que interagem com o sistema mais amiúde, inserindo novos dados ou atualizando dados já existentes. Ambos utilizam o sistema diariamente para atualizar dados esportivos e criar matérias.

## 4 Modelos de ontologias

Este capítulo é dedicado aos modelos construídos durante a pesquisa. É apresentada a ontologia de futebol e os modelos de estereótipos e esquemas narrativos desenvolvidos. A ontologia de futebol foi desenvolvida com o propósito de descrever, de forma semanticamente correta, um jogo de futebol através do mapeamento dos agentes, papéis desempenhados pelos agentes, eventos e participações possíveis. Os modelos de estereótipos e esquemas narrativos foram criados com dois propósitos, mapear situações de jogos relevantes e mapear quais informações são adequadas para cada *template* narrativo. Os modelos de estereótipos devem consultar a ontologia de futebol para obter as informações necessárias. Também são descritas as metodologias usadas na construção desses artefatos.

### 4.1 Descrição da metodologia de modelagem usada

Como o principal objetivo do sistema é dar suporte a construção de narrativas acerca de jogos de futebol, foi fundamental entender dois aspectos distintos e complementares.

O primeiro aspecto descreve os eventos, agentes e participações possíveis em um jogo de futebol, deixando claro que tipos de eventos podem ocorrer num jogo, as relações entre eles e como é a participação dos atletas nos mesmos. Para compreensão deste aspecto foi importante a participação de especialistas em futebol para orientar sobre as regras e como é a dinâmica do jogo.

O segundo identifica quais são as narrativas que podem ser adequadas para um jogo de futebol de acordo com os seus acontecimentos. A participação de um especialista em conteúdo sobre futebol para orientar quais são os padrões de "histórias e personagens" importantes foi fundamental para o entendimento deste aspecto. Assim como para definir quais informações são úteis para descrever essas "histórias e personagens".

As metodologias usadas para modelagem de cada aspecto são distintas e serão descritas, separadamente, a seguir.

#### **4.1.1 Construção de um modelo de jogo de futebol**

A construção do modelo de jogo de futebol teve como foco principal mapear os principais agentes, eventos e participações de um jogo de futebol e as relações entre eles. A participação de especialistas neste esporte foi fundamental, dentre outros motivos, porque o modelo deve respeitar as regras do jogo, assim como, responder as questões de competência identificadas. As três principais etapas da construção do modelo foram: levantamento das questões de competência, modelagem conceitual e validação do modelo proposto.

O levantamento das questões de competência foi a primeira etapa do processo de construção do modelo. As questões de competência estão relacionadas a três objetivos de negócio: Dar suporte ao modelo de estereótipos e esquemas narrativos provendo condições de implementar as regras de detecção dos estereótipos e reunir as informações necessárias para os esquemas narrativos projetados; suportar o modelo que já existia e está em produção na organização onde esta pesquisa foi realizada; implementar as melhorias e correções necessárias para o modelo atual.

Para garantir que o modelo a ser projetado responde às questões que o modelo existente responde, foi necessário estudar a base de dados existente. Foi fundamental a participação de analistas de sistema que conhecem o modelo atual. Foram realizadas algumas entrevistas com esses analistas, que ajudaram a fazer um levantamento das aplicações que usam o modelo atual. Isso foi indispensável para mapear os requisitos que o novo modelo deve continuar atendendo. Por exemplo, o modelo distinguia vinte e três tipos de eventos que podem ocorrer num jogo de futebol, e portanto o modelo proposto também precisa representar, de alguma forma, esses vinte e três tipos de eventos. O Apêndice A apresenta a relação de tipos de eventos provenientes do modelo existente.

O segundo levantamento importante, que contou com a participação de gerentes de produto, serviu para capturar as principais limitações do modelo existente. Especialistas em produto apresentaram os objetivos estratégicos para seus produtos, e a partir destes objetivos, foram levantados os requisitos que um modelo ideal precisa atender. Foi criada uma lista de novos requisitos que o modelo atual não atende mas que o modelo proposto precisa atender.

O último levantamento feito, nessa etapa de levantamento das questões de competência, foi focado em garantir que os dados e informações necessárias para os modelos de estereótipos e esquemas narrativos sejam contemplados no modelo proposto.

A segunda etapa do processo de construção do modelo foi a modelagem conceitual, propriamente dita. Foi utilizado o levantamento das questões de competência que o modelo proposto deve ser capaz de responder, realizado na etapa anterior. A modelagem foi feita de forma incremental, cada aspecto relacionado aos eventos e participações foi acoplado ao modelo separadamente, o que resultou em refatoramentos sucessivos do mesmo até a conclusão da modelagem de todos os aspectos necessários. Foi utilizada uma ontologia de fundamentação particular denominada *UFO (Unified Foundational Ontology)* que serviu como um “gabarito” durante o processo de modelagem. A seção 2.1.3 *UFO (Unified Foundational Ontology)* descreve partes dessa ontologia e apresenta trabalhos onde a mesma foi detalhadamente descrita.

A seguir descrevemos as etapas da construção do modelo proposto, ressaltando os principais passos e decisões de modelagem. A primeira versão do modelo foi desenvolvida com foco em dois objetivos importantes.

O primeiro era descrever semanticamente as participações possíveis num jogo de futebol. Participações de equipes, dos integrantes das equipes e dos árbitros, inclusive, detalhando tipos de participação diferentes entre atletas que atuam como goleiro e atletas que atuam como “jogadores de linha”. Essa distinção foi necessária devido às próprias regras do jogo que tratam esses dois tipos de participação diferentemente. Na seção 2.1.3 *UFO (Unified Foundational Ontology)* são descritos alguns conceitos relacionados a diferenciação de papéis em eventos.

Dois tipos básicos de lance foram criados para classificar os lances: lance individual e lance complexo. O lance individual é composto por apenas uma participação, por exemplo, chute para fora e o lance complexo é composto por duas participações. Por exemplo, uma defesa difícil tem duas participações: participação como atacante, que pode ser de qualquer jogador, e participação como defensor, que só pode ser de um goleiro. Outro aspecto modelado diz respeito às relações de *dependência ontológica* entre os eventos. Por exemplo, a participação de uma

equipe num jogo é existencialmente dependente da equipe, ao mesmo tempo que, a participação de um atleta no jogo é parte da participação de sua equipe no mesmo. A *UFO* considera as Relações entre intervalos de Allen para o caso de estruturas ordenadas. Veja na seção 2.1.3 *UFO (Unified Foundational Ontology)*.

Após a modelagem do jogo em si, foi necessário tratar do conceito de competição. Até então, o modelo representava com detalhes os eventos e participações ocorridos durante o jogo. No entanto, um jogo de futebol, na maioria das vezes faz parte de uma competição. As competições podem ter diferentes formatos de funcionamento. Algumas são eliminatórias, jogo a jogo, outras são baseadas em pontuação acumulada. A quantidade de fases de uma competição também varia. Enfim, foi necessário ter um modelo que represente conceitos como fase, rodada e tipo de competição. Na etapa de pesquisa antes do início da modelagem de jogo foi identificado um modelo, feito pela BBC, que descreve competições de um modo geral. Este modelo foi reutilizado para esta parte importante que faltava ser modelada. A versão mais atual da ontologia de esportes da BBC está acessível em: <http://www.bbc.co.uk/ontologies/sport/2011-02-17.shtml>.

Após a modelagem detalhada dos eventos, fases e participações de um jogo de futebol, foi feito um alinhamento entre a ontologia de domínio proposta e a ontologia de fundamentação usada a fim de validar a qualidade do modelo.

O modelo final é descrito, detalhadamente, na seção 4.2 Ontologia do Futebol.

#### **4.1.2 Definição dos estereótipos e esquemas narrativos**

Antes do processo de identificação de histórias e personagens interessantes foi feito um exercício para identificar padrões narrativos. Dezenas de crônicas de jogos foram lidas e analisadas a fim de encontrar fatos e personagens importantes comumente destacados. Foi possível notar reincidência de alguns personagens nas crônicas, como por exemplo, o artilheiro, o time mandante, o garçom, o estreante, a muralha, dentre outros. Também foram identificados fatos importantes, por exemplo, gols, substituições, pênaltis perdidos, defesas difíceis, etc, além de “estereótipos” de jogo, por exemplo, a goleada, o jogo truncado, dentre outros. Essa etapa contribuiu para reforçar a ideia de que podem existir padrões narrativos interessantes para conteúdos sobre jogos de futebol.



A segunda etapa foi uma análise estatística básica sobre eventos ocorridos em jogos de futebol.

Foi utilizada uma base de dados com jogos, atletas, escalações e estatísticas de eventos dos jogos. Existem vários tipos de eventos mapeados nessa base de dados. Gol, assistência, defesa difícil, finalização na trave, falta, cartão amarelo e cartão vermelho são alguns exemplos de tipo de evento. Com base nos eventos existentes na base de dados foi feito um levantamento das estatísticas relevantes junto aos especialistas envolvidos. Partindo dessas informações foi possível iniciar um processo comparativo de análise de dados numéricos versus análise de texto.

Foram feitas diversas consultas à base de dados para buscar estatísticas interessantes sobre os jogos. Sempre que um jogo apresentava alguma estatística interessante, era feita uma busca na web por matérias sobre o jogo para verificar se na narrativa do jogo continha alguma informação relacionada à estatística encontrada. Por exemplo, ao consultar a base de dados de jogos do campeonato brasileiro de 2010, foi identificado um jogo com uma grande quantidade de faltas e cartões, bem acima da média do campeonato. Foi o jogo com o maior número de cartões da competição. Foram encontradas várias matérias de diferentes fontes sobre este jogo na web. Uma delas destacou a quantidade elevada de cartões do jogo. Uma segunda matéria descreveu o jogo como sendo o mais violento do campeonato. Sempre que no texto de uma matéria os números que haviam sido obtidos nas consultas à bases de dados, de alguma forma eram mencionados, caracterizava-se um indício de que aqueles números eram interessantes. Quando não era possível encontrar nenhuma matéria que retratasse uma determinada estatística, era feita uma análise junto aos editores de esportes. Em algumas situações a conclusão foi que aquela estatística não era tão interessante. Em outras a conclusão foi que a informação deve ter passado despercebida pelos redatores das matérias, que teriam utilizado a mesma devido a sua importância. Como exemplo desse último cenário podemos citar um jogo da 30<sup>a</sup> rodada que foi o jogo com o maior número de faltas da competição mas não foi encontrada nenhuma matéria informando isso.

Ao final dessa etapa, foi produzido um documento com uma relação de estereótipos de jogo e personagem.

Utilizando como base o documento produzido na etapa anterior, a terceira etapa foi dedicada à definição das regras de detecção dos estereótipos. Foi necessário entender e mapear as características que um jogo deve ter para se enquadrar num estereótipo de jogo conhecido. Também foi necessário entender e mapear as características da atuação de um jogador em um jogo que o enquadram num estereótipo de personagem conhecido.

Para alguns estereótipos foi possível utilizar definições bem objetivas. Por exemplo, o estereótipo do jogo com o maior número de passes errados. Para esses casos as regras de detecção ficaram bem simples, no exemplo, a regra seria apenas verificar se a quantidade de passes erradas no jogo é maior do que as quantidades de passes errados de todos os jogos do campeonato até então. Outros estereótipos são mais complexos porque não se baseiam numa única estatística. Eles são definidos por um conjunto de informações, e cada informação pode ter um peso diferente. Por exemplo, o estereótipo de “jogão” (ou jogo movimentado). Esse estereótipo não enquadra apenas jogos com um número elevado de gols. É possível que um jogo com número de gols acima da média do campeonato não se enquadre nesse estereótipo. Para o estereótipo de “jogão” foram considerados diferentes tipos de eventos do jogo: gols, defesas difíceis, finalizações na trave e passes errados. Foi criado um índice composto pelas quantidades de ocorrências desses eventos, usando pesos diferentes para cada tipo de evento e com base nas quantidades de ocorrências desses tipos de eventos nos jogos da competição. Um índice comparativo para os jogos da competição. Os jogos com os maiores valores desse índice são os que se enquadram no estereótipo de jogão.

Foi feito um estudo empírico durante 10 rodadas do campeonato brasileiro 2011 a fim de verificar se os estereótipos detectados a partir da aplicação das regras eram coerentes. Rodada a rodada, os estereótipos detectados eram confrontados com a percepção dos especialistas envolvidos e a repercussão na mídia.

Na seção 4.3 estão listados todos os estereótipos de jogos e personagens, assim como as regras de detecção dos mesmos.

A quarta e última etapa da definição de estereótipos e esquemas narrativos serviu para se definir qual esquema narrativo seria usado para cada estereótipo. Em outras palavras, quais informações são relevantes e devem ser usadas para apre-

sentar os estereótipos detectados. Foi feito um levantamento em conjunto com os editores envolvidos para entender quais as melhores informações para contextualizar os estereótipos. Veja abaixo dois exemplos:

- 1) Para a apresentação de um jogador existe um conjunto de informações importantes: quantidade de gols, finalizações na trave e defesas difíceis de cada time. Assim como a média de gols e finalizações por partida no campeonato. Essas informações devem compor o *template* de jogador para enriquecer o conteúdo e facilitar o entendimento por parte do leitor que poderá comparar as quantidades de gols e finalizações do jogo com a média do campeonato.
- 2) A apresentação de uma goleada deve trazer outro conjunto de informações. Para esse *template* as informações usadas são: a última goleada entre os dois times que jogaram. Assim como uma lista com as maiores goleadas do campeonato. Essas informações ajudam o leitor a se aprofundar no conteúdo e fazer comparações entre diferentes goleadas.

Na seção 4.3 são mostrados os esquemas narrativos usados para cada estereótipo de jogo e personagem.

## 4.2 Ontologia do Futebol

Nesta seção, a ontologia de futebol proposta neste trabalho será explicada, detalhadamente. Como o modelo proposto é grande demais para ser visualizado por inteiro, para melhorar a visualização, o mesmo foi separado em diferentes fragmentos. Cada fragmento do modelo é detalhado através de uma imagem, uma descrição e outra imagem que mapeia os elementos daquele fragmento com a ontologia de fundamentação usada como referencia, no caso a UFO.

Esta ontologia tem como objetivo principal prover semântica de mundo real sobre o domínio de futebol. Mais especificamente sobre um jogo de futebol. Mapeando os eventos, agentes e participações possíveis. O levantamento das questões de competência descrito anteriormente neste trabalho serviu para definir quais perguntas a ontologia deverá ser capaz de responder.

### 4.2.1 Fragmento das participações em um jogo

O fragmento das participações em um jogo descreve os papéis exercidos pelos agentes que atuam no jogo. Alguns papéis que não constam nas questões de competência levantadas não estão representados, por exemplo, o papel do técnico.

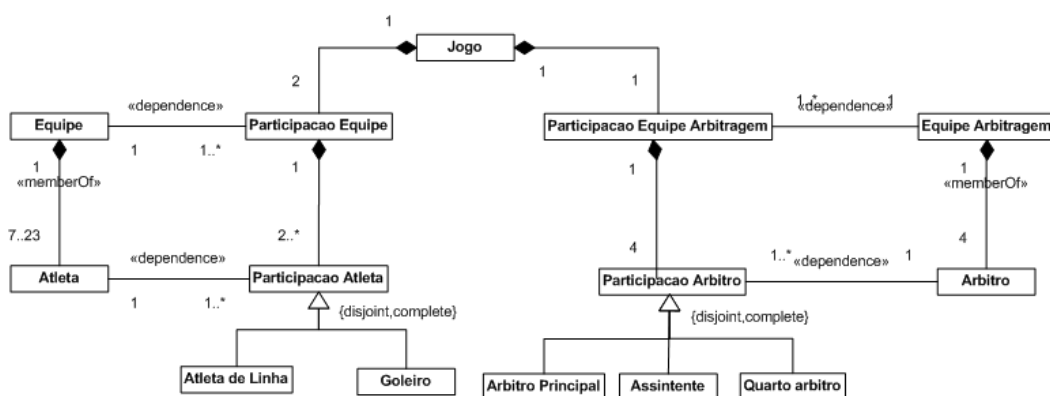


Figura 11 - Fragmento das participações em um jogo.

Em um jogo de futebol existem duas participações de equipe, e uma participação de equipe de arbitragem.

Uma *participação de equipe* é parte do evento *jogo* e é existencialmente dependente do objeto *equipe*. Ao mesmo tempo, uma *participação de atleta* é parte de uma *participação de equipe*, e por transitividade, parte do evento *jogo*. Além de ser existencialmente dependente do objeto *atleta*. Para isso ser verdade, o *atleta* deve fazer parte da *equipe*. O modelo suporta dois tipos distintos de *participação de atleta*, *participação como goleiro* e *participação como atleta de linha*. Esses são os únicos tipos de *participação de atleta* e um objeto *atleta* só pode ter um desses tipos numa participação. Acontece de um atleta ter no mesmo jogo duas participações distintas, uma de cada tipo. Desta forma, foi necessário criar uma partição disjunta e completa para representar corretamente os tipos de participação de atleta. Essa distinção foi necessária para representar as participações de atletas em lances do jogo. Na seção 4.2.5 Fragmento dos Lances Complexos as participações de atletas em lances do jogo são exibidas detalhadamente.

Analogamente, uma *participação de equipe de arbitragem* também é parte do evento *jogo* e é existencialmente dependente do objeto *equipe de arbitragem*. Da mesma forma que uma *participação de árbitro* é parte de uma *participação de equipe de arbitragem*, e por transitividade, parte do evento *jogo*. Também é existencialmente dependente do objeto *árbitro*. Logo, o *árbitro* deve fazer parte da *equipe de arbitragem*. O modelo suporta três tipos distintos de *participação de árbitro*, *participação árbitro principal*, *participação como assistente* e *participação com quarto árbitro*. O modelo não relaciona os tipos de participação de árbitro com os lances do jogo, apenas captura a formação da equipe de arbitragem de um jogo.

A Figura 12 exhibe este fragmento da ontologia alinhado com a *UFO*.

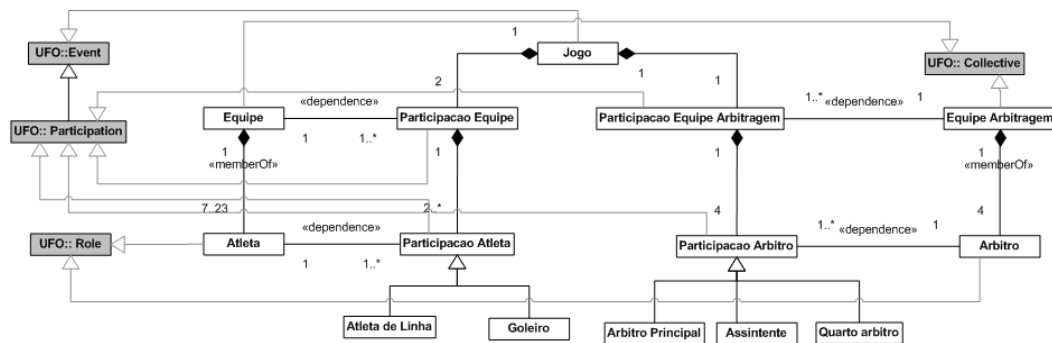


Figura 12 - Fragmento das participações em um jogo alinhado com a UFO.

Vale ressaltar que a modelagem adotada permite a distinção entre os dois tipos de participação de atleta. Ao considerar a participação de atleta como *Participation*, estamos considerando que a mesma é um *Event*. Participações são especializações de eventos e por isso herdam as propriedades de eventos. De acordo com a *UFO* todo evento tem uma duração, ou para ser mais preciso, uma extensão temporal. Dessa forma é possível criar axiomas que restringem que um atleta pode ter tipos de participação diferentes no mesmo jogo desde que as mesmas aconteçam durante intervalos temporais que não se intersectam. Isso reproduz, fielmente, as regras do jogo. O conceito que foi usado para mapear as relações entre intervalos temporais é baseado nas Relações entre intervalos de Allen[Allen 1983] que foram apresentadas de forma simplificada na seção 2.1.3.

### 4.2.2 Fragmento das etapas de disputa de um jogo

Este fragmento tem como principal função mapear as etapas de um jogo de futebol e suas relações com os lances de jogo que podem acontecer durante as etapas. Existem tipos de lance que não podem ocorrer em qualquer etapa do jogo.

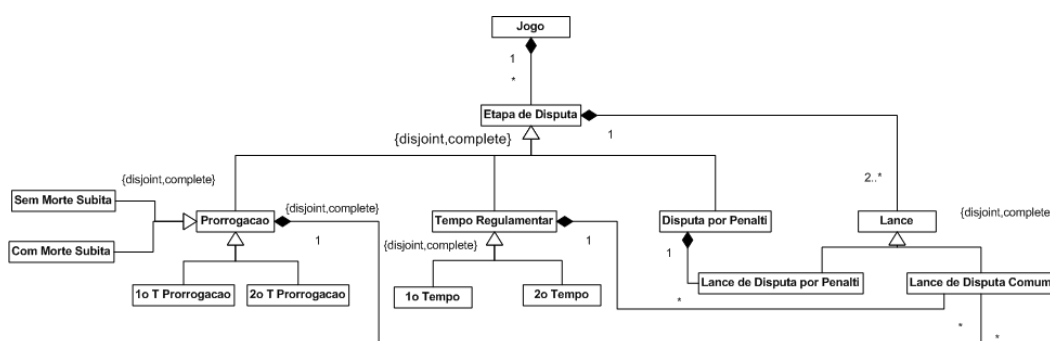


Figura 13 - Fragmento das etapas de disputa de um jogo.

Todo lance de um jogo ocorre em uma *Etapa de jogo*. Um jogo é composto por uma ou mais etapas. Isso varia de acordo com as regras da competição a qual o jogo faz parte. Existem três tipos de etapas de jogo: *Tempo Regularizar*, *Prorrogação* e *Disputa por Pênaltis*. Esses são todos os tipos de etapa que um jogo pode ter, mais uma vez foi criado uma partição disjunta e completa para representar corretamente os tipos.

As etapas *Tempo Regularizar* e *Prorrogação* são divididas em primeiro tempo e segundo tempo. Já para a etapa *Disputa por Pênalti* não existe essa distinção, mas sim um conjunto de lances específicos que ocorrem durante esta etapa. No caso, as cobranças de pênaltis, alternadas entre os times. Para a etapa de *Prorrogação* além de uma partição disjunta e completa para distinguir primeiro tempo e segundo tempo, existe uma segunda, com as mesmas características, para distinguir se a etapa é *Com Morte Súbita* ou *Sem Morte Súbita*. Uma *Prorrogação Com Morte Súbita* significa que o jogo termina caso alguns dos times faça um gol, sem a necessidade de cumprir a duração total prevista para a etapa.

Foram criadas duas especializações de lance para mapear quais tipos de lance podem ocorrer em quais etapas de jogo. *Lance de Disputa Comum* é um tipo de lance que pode acontecer tanto na etapa *Tempo Regularizar* quanto na etapa

*Prorrogação*. Lances do tipo *Lance de Disputa por Pênalti* só podem acontecer na fase *Disputa por Pênalti*. Pode parecer que cobrança de pênalti é um tipo de lance que pode acontecer em qualquer fase. Porém, existem diferenças entre cobrança de pênalti durante as fases *Tempo Regular* e *Prorrogação* e cobrança de pênalti durante a fase *Disputa por Pênalti*. Nesta última, uma cobrança de pênalti não pode originar um passe, além de não ser permitido o rebote após a defesa do goleiro. Nas outras duas é permitido que tanto uma cobrança de pênalti seja um passe quanto que uma defesa de uma cobrança gere um rebote.

Veja abaixo como o fragmento apresentado estendeu alguns conceitos da *UFO*.

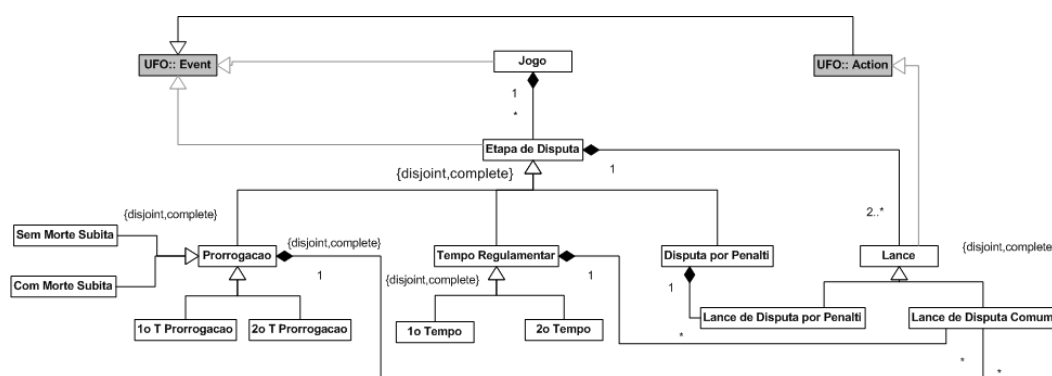


Figura 14 - Fragmento das etapas de disputa de um jogo alinhado com a UFO.

Os eventos podem ser atômicos ou complexos, dependendo de sua estrutura mereológica, ou seja, enquanto eventos atômicos não têm partes impróprias, eventos complexos são agregações de pelo menos dois eventos (que podem ser eles próprios atômicos ou complexos) [Guizzardi et al, 2008]. Podemos considerar que um jogo é um evento complexo composto de outros eventos que são as etapas de disputa. As etapas de disputa também são eventos complexos, compostos de lances de jogo (ex.: gol, cartão amarelo, falta, defesas). Pode-se dizer que os lances de jogo são ações. Ações são eventos intencionais realizados por agentes participantes.

#### 4.2.3 Fragmento do lance de disputa comum

*Lance de Disputa Comum* são eventos que podem ocorrer em todas as etapas de disputa exceto na etapa *Disputa por Pênalti*. Todos os jogos tem pelo menos

uma etapa de disputa em que esses eventos acontecem. A grande maioria dos eventos mapeados no modelo fazem parte dessa especialização.

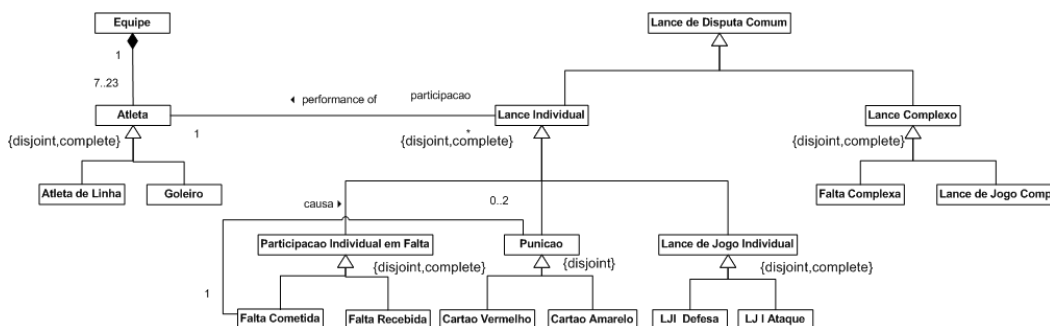


Figura 15 - Fragmento do lance de disputa comum.

Mais uma vez, utilizou-se no modelo, a estrutura mereológica de eventos, proposta pela *UFO*. *Lance de Disputa Comum* recebeu duas especializações. *Lance individual* que tem participação de apenas um jogador e *Lance Complexo* composto duas instâncias de *Lance Individual*, cada uma com participação de jogadores diferentes. *Lance individual* é um evento atômico e *Lance complexo* é um evento complexo, formado pela agregação de dois lances atômicos.

Um *Lance Individual* pode ser uma *Participação Individual em Falta*, uma *Punição* ou um *Lance de Jogo Individual*. Uma *Participação Individual em Falta* pode ser uma *Falta Recebida* ou uma *Falta Cometida*. A *Punição*, pode ser um cartão amarelo ou cartão vermelho. Vale notar que existe uma relação de causalidade entre esses eventos, ou seja, uma *Punição* é causada por uma *Participação Individual em Falta*. Em outras palavras, só existe uma *Punição* a partir da existência de uma *Participação Individual em Falta*. A terceira especialização de *Lance Individual* abrange a grande maioria dos eventos atômicos que ocorrem em um jogo. *Lance de Jogo Individual* abrange todos os eventos comuns do jogo, ou seja, que acontecem em situações normais de prática dos fundamentos do esporte. Chutes, passes, cobranças de falta, etc. Para representar esse conjunto de eventos foi criada uma partição disjunta e completa para classificá-los entre *LJI Defesa* (*Lance de Jogo Individual de Defesa*) e *LJI Ataque* (*Lance de Jogo Individual de Ataque*).



*Lance Complexo* é o conjunto de lances caracterizados pela combinação de dois lances individuais. Tem duas especializações. *Falta Complexa* são os lances faltosos que envolvem duas participações atômicas. Uma falta onde um jogador usa excesso de força contra um jogador do time adversário é um exemplo. Existem duas participações atômicas, uma de quem comete a falta e outra de quem sofre. Já uma falta por colocar a mão na bola não é considerada *Falta Complexa* por ter somente uma participação. A segunda especialização de *Lance Complexo* é chamada de *Lance de Jogo Complexo*, e compreende os lances comuns (não faltosos) formados por dois lances do tipo *Lance de Jogo Individual*. Como exemplo temos *Finalização Defendida* que é um evento complexo que agrega dois eventos atômicos:

- 1) Uma finalização não convertida, com *participação de atleta*. Qualquer atleta pode realizar uma finalização contra o gol do adversário
- 2) E uma defesa, com *participação de atleta goleiro*. Somente atletas que tem participação como goleiro podem realizar defesas.

A relação entre *Lance Complexo* e *Lance Individual* é descrita detalhadamente na seção 4.2.5 Fragmento dos Lances Complexos.

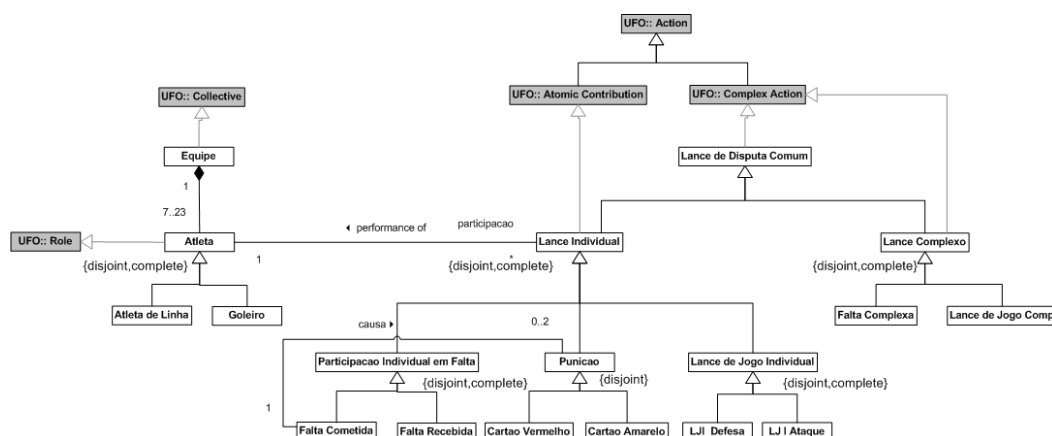


Figura 16 - Fragmento do lance de disputa comum alinhado com a UFO.

#### 4.2.4 Fragmento do lance de jogo individual

*fesa*, que só acontece em *participação de atleta goleiro*. Na seção 4.2.5 Fragmento dos Lances Complexos são exibidas as relações entre as participações de atleta e os lances de jogo.

*Lance de Jogo Individual de Defesa* tem como especializações, *Roubada de bola*, *Bloqueio*, *Defesa* e *Gol Contra*. *Roubada de Bola* e *Gol Contra* são eventos atômicos que não fazem parte de nenhum evento complexo (lance complexo). Já *Defesa* e *Bloqueio* são eventos, que fazem parte dos eventos complexos *Finalização Bloqueada* e *Finalização Defendida*, respectivamente. O modelo captura a informação de quais *Defesas* foram consideradas difíceis. Para isso, existe uma especialização de *Defesa*, chamada *Defesa difícil*.

*Lance de Jogo Individual de Ataque* tem uma quantidade maior de níveis de especializações. O primeiro nível cria uma partição disjunta e completa com três especializações: *Posicionamento Impedido*, *Recebimento de Passe* e *Ação Ofensiva*. *Posicionamento Impedido* é quando o posicionamento de um atleta do time que tem a posse de bola, no momento em que outro atleta do mesmo time realiza um passe, caracteriza uma posição de impedimento. *Recebimento de passe* é a participação de um jogador ao receber um passe feito por outro jogador da mesma equipe. *Ação ofensiva* são ações feitas por jogadores que estão com a posse de bola. Este tipo de evento, abrange um grande número de eventos atômicos, e tem duas partições ortogonais.

Uma delas existe para capturar o estado original da bola no momento do evento. Completa, disjunta e com duas especializações: *Ação com Bola Rolando*, que representa as ações realizadas com a bola em movimento, e *Cobrança*, que representa as ações realizadas com a bola parada. *Cobrança* também tem especializações que detalham o tipo de cobrança., que pode ser *Cobrança de escanteio*, *Cobrança de lateral* e *Cobrança de falta*. Esses são os tipos mapeados no levantamento das questões de competência mas outros tipos podem existir. Um exemplo de cobrança não mapeado nas questões de competência mas é uma especialização de cobrança é Cobrança de tiro de meta. Por isso o *generalization set* não foi indicado como completo.

A outra partição de *Lance de Jogo Individual de Ataque*, está relacionada à intenção do atleta. Esta partição é completa e disjunta com duas especializações:

*Passe*, que representa as ações originadas da intenção de um jogador passar a bola para outro jogador da mesma equipe, e *Finalização*, que representa as ações originadas da intenção de um jogador realizar uma finalização contra o gol da equipe adversária.

Os tipos de *Passe* mapeados no modelo são: *Passe Errado*, quando um jogador que realiza um passe não consegue fazê-lo de forma correta e com isso a bola não chega até o jogador para o qual o passe era destinado; *Passe comum*, quando um jogador que realiza o passe consegue fazê-lo de forma correta e com isso a bola chega até o jogador para o qual o passe era destinado; *Assistência*, quando um jogador realiza um passe correto seguido de um gol feito pelo jogador que recebeu o passe; *Passe Impedido*, quando o jogador realiza um passe que é invalidado por conta da posição irregular de impedimento de quem recebeu o passe. Das quatro especializações descritas, apenas uma não compõe uma ação complexa, *Passe Errado*. As outras três podem ser consideradas *Atomic Actions* que compõem ações complexas. Todas serão novamente abordadas na seção 4.2.5 Fragmento dos Lances Complexos.

Para os eventos *Finalização*, novamente, foi feito um *generalization set*. Dessa vez para representar o resultado da finalização. Quatro especializações cobrem todas as variações desse quesito mapeadas no levantamento das questões de competência. *F Trave* (*Finalização na Trave*), *F Fora* (*Finalização para Fora*), *F Gol Intercep* (*Finalização no Gol Interceptada*) e *F Gol Convert* (*Finalização no gol Convertida*). *Finalização na Trave* e *Finalização para Fora* são eventos atômicos, que não compõem nenhum evento complexo. As outras duas especializações são *Atomic Actions* que compõem ações complexas e serão, novamente, abordadas na próxima seção.

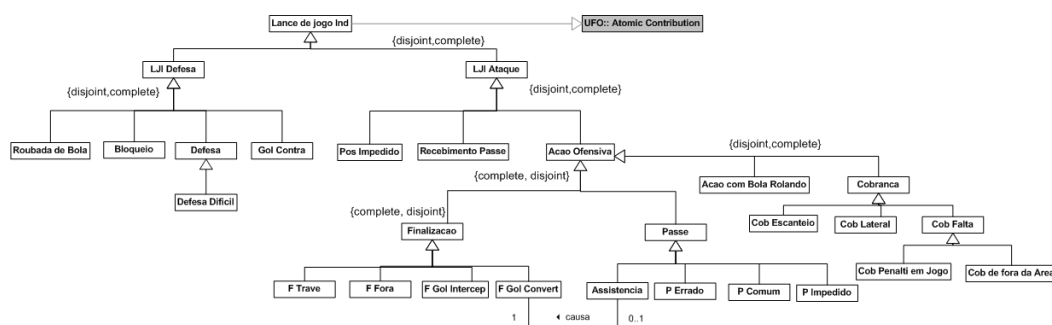


Figura 18 - Fragmento do lance de jogo individual alinhado com a UFO.

Neste fragmento são apresentadas diversas especializações de *Lance de Jogo Individual*, que é uma *Atomic Action*. Por consequência, todas as especializações também são *Atomic Action* com características adicionais específicas de cada especialização.

#### 4.2.5 Fragmento dos Lances Complexos

O fragmento de lance complexo foi brevemente definido na seção 3.2.3) *Fragmento do lance de disputa comum*. Nas próximas seções serão mostradas as interações que caracterizam cada evento complexo. Com indicação dos eventos atômicos e das participações envolvidos no mesmo. Foram criados dois conjuntos de interações, aqui denominados, Interações de lances faltosos e Interações de lances de jogo.

##### 4.2.5.1 Lances faltosos

Esta seção é dedicada ao detalhamento das partes do modelo que representam as faltas de um jogo de futebol. Faltas são ações não permitidas durante o jogo de acordo com as regras.

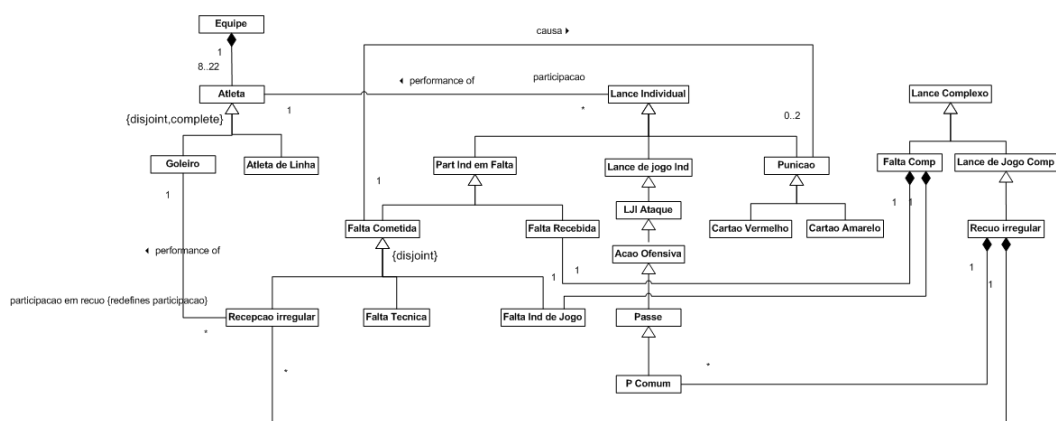


Figura 19 – Fragmento do lances faltosos.

Todas as situações de falta do jogo estão relacionadas a uma *Part Ind em Falta* (*Participação Individual em Falta*). Esta especialização de *Lance Individual* tem duas especializações representadas por um *generalization set* completo e disjunto com duas partições: *Falta Cometida* e *Falta Recebida*. Essa distinção permite saber, exatamente, quais jogadores sofreram faltas e quais cometeram faltas. Vale notar que, de um modo geral, qualquer jogador (atleta) pode ter participação tanto em *Falta Cometida* quanto em *Falta Recebida*, independente do tipo de participação do jogador (*Goleiro* ou *Atleta de Linha*). Foi mapeado apenas um tipo de falta que se aplica exclusivamente a participações como *Goleiro*.

Ao todo, três tipos diferentes de falta estão representados no modelo. Qualquer falta cometida pode gerar uma punição, porém, o modelo permitiria que apenas tipos específicos de falta gerem punição devido às especializações criadas para *Falta Cometida*. *Punição* no modelo proposto são as advertências sancionadas com cartão durante um jogo de futebol.

*Recepção irregular* é um tipo de falta que acontece apenas com goleiros. É quando o goleiro recebe com as mãos, dentro da sua área, uma bolada recuada por outro jogador da sua equipe. Por isso no modelo existe uma relação chamada *participacao em recuo* entre a participação *Goleiro* e a ação *Recepção Irregular*. Caso o recuo seja feito com a cabeça ou com o peito, não será caracterizado *Recepção Irregular*. Para estes tipo de recuo, a recepção com as mãos por parte do goleiro é permitida.

*Falta tecnica* (*Falta Técnica*) é uma especialização de *Falta Cometida* que não faz parte de nenhum evento complexo. Este tipo conta com participação de apenas

um jogador. Se enquadram neste tipo as faltas oriundas de situações como: reclamação, mão na bola e ofensas feitas ao árbitro.

A última situação de falta mapeada no modelo é a *Falta Comp (Falta Complexa)*. É o tipo de falta que, normalmente, ocorre com maior frequência. É uma ação complexa formada por uma *Atomic Action* chamada *Falta Ind de jogo (Falta Individual de jogo)* e um evento atômico chamado *Falta Recebida*. Temos como exemplos de *Falta Complexa* quando um jogador segura, derruba, usa de excesso de força na marcação ou agride um jogador do time adversário. O modelo permite então mapear para cada falta as duas participações de quem comete e de quem sofre a falta.

A figura a seguir exibe o alinhamento do fragmento do modelo apresentado nesta seção com a *UFO*.

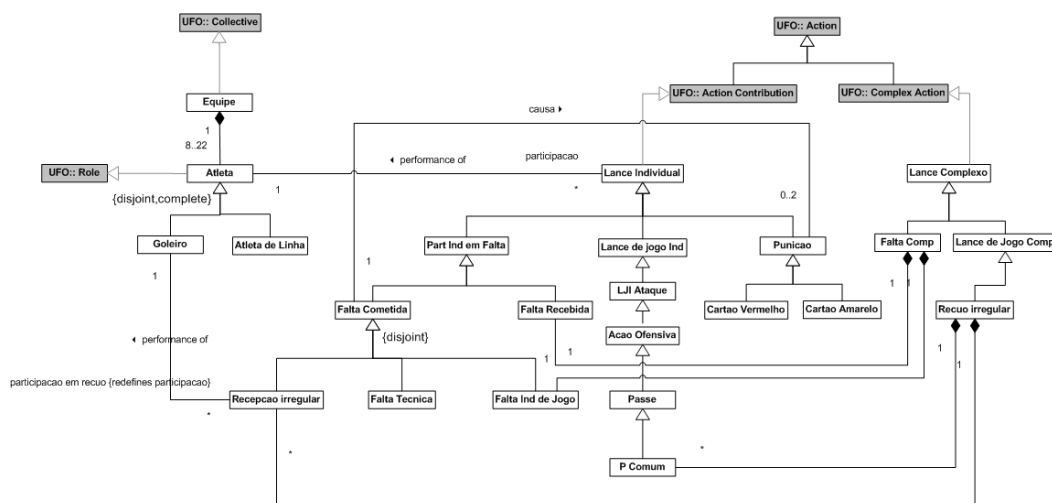


Figura 20 – Fragmento dos lances faltosos alinhado com a UFO.

#### 4.2.5.2 Lances de jogo

As ações que envolvem as práticas dos fundamentos sem infração de regras, aqui denominadas, lances de jogo, serão detalhadamente descritas nesta seção.

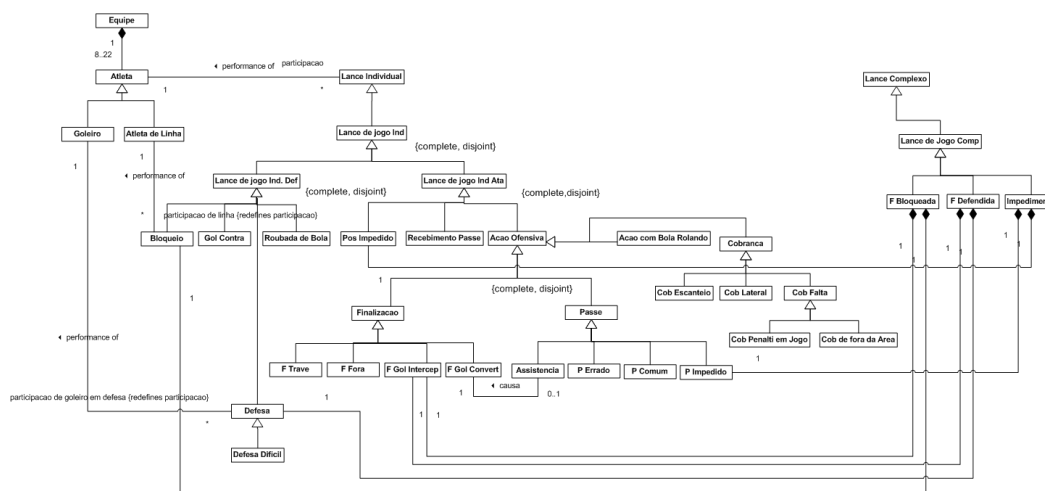


Figura 21 – Fragmento do lances válidos.

*Lance de jogo Ind (Lance de Jogo Individual)* e *Lance de Jogo Comp (Lance de Jogo Complexo)*, representam da forma mais geral, respectivamente, os eventos atômicos e eventos complexos que constituem o conjunto de lances válidos .

Foi criado um *generalization set* disjuncto e completo para os eventos atômicos (*Lance de Jogo Individual*) com duas partições: *Jance de Jogo Ind Def (Lance de jogo Individual de Defesa)* e *Jance de Jogo Ind Ata (Lance de jogo Individual de Ataque)*.

Para os eventos complexos estas partições não fazem sentido porque os eventos complexos são formados por dois eventos atômicos, sendo um de ataque e outro de defesa.

Existem quarenta e sete tipos de lance válido previstos no modelo. Distribuídos entre diversos fragmentos. Esta conta pode ser feita somando:

- Quantidade de tipos de evento atômico de defesa (especializações de *Lance de jogo Individual de Defesa*). *Gol Contra* e *Roubada de Bola*.
- Quantidade de tipos de evento atômico de ataque (especializações de *Lance de jogo Individual de Ataque*). Finalização na trave, Finalização para fora, Finalização para o gol convertida, Assistência, Passe Errado, Passe Comum e Passe Impedido. Somando os tipos citados dá um total de oito. Porém, ao analisar o elemento *Ação Ofensiva* do modelo é notável um segundo aspecto de especialização, no caso, o do estado da bo-



la no momento do evento. Com cinco possibilidades aplicáveis aos oito tipos de eventos atômicos de ataque, pelo produto cartesiano, o total de tipos passa para quarenta. Afinal, de acordo com o modelo, uma *Finalização na trave* pode ser com bola rolando, por cobrança de escanteio, cobrança de lateral, cobrança de falta dentro da área ou cobrança de falta de fora da área.

Quantidade de tipos de evento complexo (especializações de *Lance de jogo complexo*). *Finalização Bloqueada* e *Finalização Defendida e Impedimento*. Como esses dois tipos de evento complexo tem um evento atômico de ataque, o segundo aspecto de especialização de Ação Ofensiva também se aplicam aos tipos de evento complexo. De forma análoga aos tipos de evento atômico de ataque, pelo produto cartesiano, o total de tipos de evento complexo passa para dez. Outro evento complexo mapeado no modelo é o Impedimento. O impedimento não é considerado uma falta, embora seja um evento determinado pela interrupção do árbitro. O *Impedimento* é um evento complexo que acontece quando dois eventos atômicos ocorrem simultaneamente:

- 1) Um jogador está em posição de impedimento. Em outras palavras, ele está mais próximo da linha do gol adversário do que a bola e o penúltimo jogador da equipe adversária. No modelo, essa situação é representada por *Pos Impedimento (Posição de Impedimento)*.
- 2) No momento em que a bola toca ou é tocada por um outro jogador de sua equipe e o jogador que está em posição de impedimento está, na opinião do árbitro, envolvido ativamente na jogada. O modelo define esta ação como *P Impedido (Passe Impedido)*.

Do ponto de vista do alinhamento com a UFO, o fragmento do modelo que representa os lances válidos é semelhante ao fragmento que representa os lances faltosos.

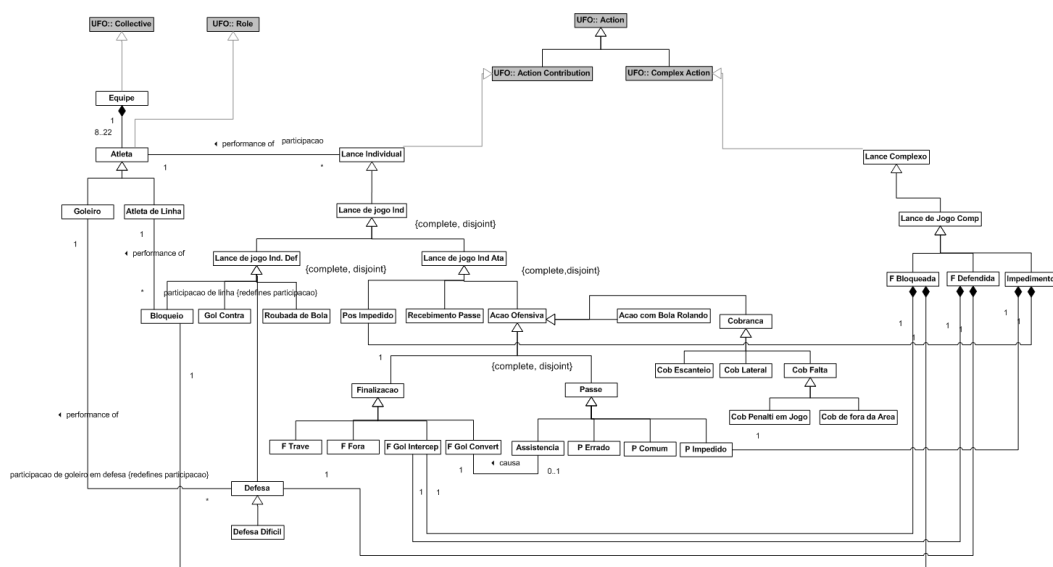


Figura 22 – Fragmento dos lances válidos alinhado com a UFO.

#### 4.2.6 Resumo do modelo

O modelo descrito nesta seção tem como principal objetivo mapear, de forma semanticamente correta, os principais eventos que ocorrem durante a partida e as relações entre os mesmos. Representar todas as regras do jogo não é o objetivo do modelo, porém, para os eventos que devem ser representados, esta representação deve considerar as regras.

Os eventos mapeados são aqueles que tem relação com os atletas participantes e suas performances nos diversos lances de jogo (faltas, finalizações, defesas, etc). Situações que podem ocorrer durante o jogo mas não tem relação direta com as participações dos atletas não constam nas questões de competências levantadas até o momento. Como exemplo dessas situações temos, invasão de campo, interrupção de jogo por falta de energia ou qualquer situação atípica que pode ocorrer durante um jogo.

Com este mapeamento é possível obter todas as informações estatísticas desejadas sobre as performances das equipes e jogadores.

As principais vantagens do modelo proposto em relação ao modelo existente na organização onde a pesquisa foi realizada são:

1. Relação entre os lances. No modelo existente é possível obter os totais de ocorrências de cada *scout* (lance de jogo) relacionados a cada participação de atleta. Mas não é possível obter as relações entre *scouts* (relação de causalidade ou precedência, por exemplo). Por exemplo, era possível saber que um jogador A teve x faltas recebidas no jogo, enquanto o jogador B teve y faltas cometidas no jogo. Mas não era possível saber quais faltas do jogador B foram cometidas no jogador A. No modelo proposto é possível instanciar todas as relações entre lances previstas nas questões de competência.
2. Diferenciação entre papéis. O modelo anterior prevê o papel Atleta mas não prevê tipos de participação de atleta (Goleiro e Atleta de Linha). Isso faz com que o modelo permita, por exemplo, que um jogador de linha realize uma defesa, o que, semanticamente, é incorreto. No modelo proposto, a diferenciação de papéis sugerida provê maior semântica de mundo real.
3. Fases do jogo e lances possíveis. O modelo proposto mapeia que eventos (tipos de) podem ocorrer durante as fases do jogo (tipos de ). Ele prevê que fase de disputa por pênalti vários lances de jogo comuns não se aplicam. Por exemplo, passe, roubada de bola ou cobrança de lateral. O modelo existente faz tais distinções.

Enfim, o modelo proposto corrige um número de problemas conceituais existentes no modelo original, obtendo-se então, uma versão mais fiel ao domínio representado. Além de mapear novas situações que o modelo original não mapeia.

Do ponto de vista da utilização da UFO no processo de modelagem, pode-se dizer que, o fragmento mais utilizado foi a UFO-B (Ontologia de eventos). Os principais conceitos abordados por essa ontologia de fundamentação utilizados foram: Diferenciação entre papéis; Eventos e Ações (estrutura mereológica de eventos – eventos atômicos ou complexos) e Relações temporais entre eventos.

### 4.3 Modelo de estereótipos e narrativas

São apresentados nesta seção todos os estereótipos implementados no protótipo e suas respectivas narrativas. Foram implementados *templates* para enriquecimento de texto, onde cada *template* representa uma narrativa de um estereótipo específico.

Os estereótipos identificados nessa pesquisa estão mapeados numa matriz com duas dimensões : tipo de estereótipo x intervalo temporal. Tipo de estereótipo diz respeito ao objeto estereotipado, no caso, um jogo ou uma atuação de um atleta em um jogo (que define o personagem dele no jogo). Intervalo temporal está relacionado ao período observado para a detecção do estereótipo.

Veja na Figura 23 - Matriz de estereótipos uma matriz com alguns estereótipos identificados no início do projeto. Aqueles em destaque foram implementados no protótipo.

	Jogo em si	Na edição do campeonato	Na história
Jogo	<ul style="list-style-type: none"> <li>- Jogo movimentado("Jogão")</li> <li>- Jogo truncado</li> <li>- Goleada</li> <li>- Placar duvidoso</li> <li>- Jogo morno("Pelada")</li> <li>- "Dramático"</li> </ul>	<ul style="list-style-type: none"> <li>- Mudança de posição na tabela</li> <li>- Sequência no campeonato</li> <li>- Maiores/Menores performances</li> <li>- Com discrepância em relação a média do campeonato</li> <li>- Final</li> </ul>	<ul style="list-style-type: none"> <li>- Maiores/Menores performances no confronto</li> <li>- Sequência no confronto</li> <li>- Última goleada no confronto</li> <li>- Maior goleada no confronto</li> </ul>
Personagem	<ul style="list-style-type: none"> <li>- Finalizador</li> <li>- Caçado</li> <li>- Caçador</li> <li>- Artilheiro</li> <li>- Garçom</li> <li>- Ladrão de bola</li> <li>- Talismã</li> </ul>	<ul style="list-style-type: none"> <li>- Maior finalizador do campeonato</li> <li>- Mais caçado do campeonato</li> <li>- Maior caçador do campeonato</li> <li>- Maior artilheiro do campeonato</li> <li>- Maior ladrão de bola do campeonato</li> <li>- Maior garçom do campeonato</li> <li>- Maior talismã do campeonato</li> </ul>	<ul style="list-style-type: none"> <li>- Talismã</li> </ul>

Figura 23 - Matriz de estereótipos

#### 4.3.1 Regras de detecção de estereótipos

As regras usadas para detecção de estereótipos foram estabelecidas através de um estudo empírico, com a ajuda de especialistas e através de análise estatística de dados de jogos de futebol. A dinâmica utilizada foi descrita na seção 3.1.1 *Definição dos estereótipos e esquemas narrativos*. Aqui serão exibidas as regras de detecção dos estereótipos implementados no protótipo. Primeiro os estereótipos de jogo e depois os estereótipos de personagem.

Estereótipo Jogo Movimentado (“Jogão”)	
Descrição	Regras
Template que inicialmente era chamado de “Jogão”. É acionado quando o jogo tem um alto índice de objetividade. O índice de objetividade de um jogo é composto por: quantidade de gols, quantidade de defesas difíceis e quantidade de finalizações na trave.	<ul style="list-style-type: none"> <li>• Jogos que estejam entre os 5% com maior índice de objetividade do campeonato</li> <li>• Jogos que não estejam entre os 5% com mais passes errados do campeonato</li> <li>• Índice de objetividade = <math>(5 * PG + 2 * PFT + 3 * PDD) / 10</math></li> <li>• Onde PG = performance em gols, PFT = performance em finalizações na trave e PDD = performance em defesas difíceis</li> <li>• Todas as variáveis de performance acima são uma proporção em relação a performance máxima ocorrida na competição</li> </ul>

Tabela 1 - Estereótipo Jogo Movimentado

O índice de objetividade do jogo, principal informação na qual a regra é baseada, sofreu algumas mudanças antes de chegar no modelo descrito. Inicialmente, o índice era composto em quarenta por cento pela quantidade de gols, trinta por cento pela quantidade de finalizações na trave e trinta por cento pela quantidade de defesas difíceis. Foi verificado que a quantidade de gols deveria ter ainda mais importância e sua contribuição para o índice aumentou de quarenta para cinquenta por cento. Ao mesmo tempo, a contribuição da quantidade de finalizações na trave para o índice foi reduzida de trinta para vinte por cento. Isto porque finalização na trave ocorre com pouca frequência e a quantidade é baixa. Dessa forma, o índice estava valorizando muito jogos com finalização na trave. O índice se mostrou mais adequado após essas alterações.

Além disso, no começo a detecção do estereótipo era baseada exclusivamente no índice de objetividade. Somente depois que o índice de passes errados foi considerado.

Veja a seguir a descrição do Estereótipo de Jogo Morno.

Estereótipo Jogo Morno ("Pelada")	
Descrição	Regras
<p><i>Template</i> que inicialmente era chamado de "Pelada". É acionado quando o jogo tem um baixo índice de objetividade. O índice de objetividade de um jogo é composto por: quantidade de gols, quantidade de defesas difíceis e quantidade de finalizações na trave.</p>	<ul style="list-style-type: none"> <li>Jogos que estejam entre os 5% com menor índice de objetividade do campeonato</li> <li>Jogos que ao mesmo tempo estejam entre 50% com menor índice de objetividade e entre os 10% com maior índice de passes errados</li> <li>Índice de objetividade = <math>(5 * PG + 2 * PFT + 3 * PDD) / 10</math></li> <li>Onde PG = performance em gols, PFT = performance em finalizações na trave e PDD = performance em defesas difíceis</li> <li>Todas as variáveis de performance acima são uma proporção em relação a performance máxima ocorrida na competição</li> </ul>

Tabela 2 - Estereótipo Jogo Morno

O estereótipo de Jogo Morno surgiu como o inverso exato do estereótipo de Jogo Movimentado, isto é, jogos que estejam entre os dez por cento com menor índice de objetividade., ou jogos que estejam entre os dez por cento com maior índice de passes errados.

No entanto, foram identificados alguns jogos que estavam entre os dez por cento com maior número de passes errados mas tinham um alto nível de objetividade. Com isso esses jogos eram enquadrados como Jogo Morno mesmo com um bom índice de objetividade. Por isso a regra foi ajustada para só detectar o estereótipo se o índice de passes errados for alto e o índice de objetividade não for alto.

O próximo estereótipo de jogo que será detalhado é o Estereótipo de Jogo truncado.

Estereótipo Jogo Truncado	
Descrição	Regras

<p>O <i>template</i> de jogo truncado é acionado para jogos que tenham um índice de truncamento alto. O índice de truncamento é composto por: quantidade de faltas, quantidade de cartões amarelos e quantidade de cartões vermelhos. Apenas jogos que tenham um ou mais desses <i>scouts</i>, significativamente, acima da média recebem esse <i>template</i> como indicação.</p>	<ul style="list-style-type: none"> <li>Jogos que estejam entre os 5% com maior índice de truncamento do campeonato</li> <li>Índice de truncamento = <math>(4 * PF + 4 * PCA + 2 * PCV) / 10</math></li> <li>Onde PF = performance em faltas, PCA = performance em cartões amarelos e PCV = performance em cartões vermelhos</li> <li>Todas as variáveis de performance acima são uma proporção em relação a performance máxima ocorrida na competição</li> </ul>
--	--

Tabela 3 - Estereótipo Jogo Truncado

Esse estereótipo tomou como base, exclusivamente, o índice de truncamento. Este índice, inicialmente, composto em quarenta por cento pela quantidade de faltas, trinta por cento pela quantidade de cartões amarelos e trinta por cento pela quantidade de cartões vermelhos.

Após algumas rodadas verificando o índice foi feito um ajuste para reduzir a contribuição da quantidade de cartões vermelhos para o índice de trinta para vinte por cento. Cartão vermelho ocorre com pouca frequência e em quantidade baixa. Dessa forma, o índice estava valorizando muito jogos com cartão vermelho e por isso foi ajustado.

Estereótipo Goleada	
Descrição	Regras
<p>O <i>template</i> de goleada é acionado sempre que o placar da partida tiver diferença maior ou igual a 3.</p>	<ul style="list-style-type: none"> <li>Jogos que tenham o placar da partida com diferença entre o placar do time mandante e o placar do time visitante maior ou igual a 3.</li> </ul>

Tabela 4 - Estereótipo Goleada

A regra de detecção do estereótipo Goleada é a mais simples de ser implementada, bastando ter a informação do jogo com o placar para calcular. Esta regra não avalia nenhum outro jogo.

Estereótipo Jogo recorde no campeonato	
Descrição	Regras
Esse <i>template</i> é acionado sempre que um jogo for o primeiro no ranking do campeonato em quantidade de ocorrências de um determinado <i>scout</i> .	<ul style="list-style-type: none"> <li>Jogos que tenham como quantidade de ocorrências de um determinado <i>scout</i> o valor máximo de ocorrências registradas numa partida do campeonato</li> <li>Considera todos os jogos realizados do campeonato até a data do jogo</li> <li>A lista completa de <i>scouts</i> existentes pode ser consultada no Apêndice A.</li> </ul>

Tabela 5 - Estereótipo Jogo Recorde no campeonato

Algumas aplicações do estereótipo Jogo recorde na edição do campeonato foram desprezadas porque não ficaram interessantes. Por exemplo, jogo com quantidade recorde de impedimentos na edição do campeonato.

Veja a seguir o Estereótipo Placar duvidoso, que foi o estereótipo que mais sofreu mudança nas regras de detecção durante o experimento.

Estereótipo "Placar duvidoso"	
Descrição	Regras
O <i>template</i> de Placar duvidoso é acionado sempre que o placar final do jogo não for condizente com os números da partida nos <i>scouts</i> de objetividade. Os <i>scouts</i> de objetividade considerados são: Gols, finalizações na trave, defesas difíceis e quantidade total de finalizações. Se um time for, significativamente, superior ao outro nesses <i>scouts</i> e es-	<ul style="list-style-type: none"> <li>Jogos que tenham <math>(IND\_F\_MAN &gt; 1.3 \cdot IND\_F\_VIS)</math> e <math>(DD\_V &gt; 1.5 \cdot DD\_M)</math> e <math>(DD\_M &gt; 1)</math> e <math>(FG\_V &gt; FG\_M)</math> e vice versa</li> <li>Jogos que tenham <math>(IND\_F\_MAN &gt; 1.5 \cdot IND\_F\_VIS)</math> e <math>(DD\_V &gt; 2 \cdot DD\_M)</math> e <math>(DD\_M &gt; 1)</math> e <math>(FG\_V</math></li> </ul>



<p>sa superioridade não for refletida no placar, o <i>template</i> de placar duvidoso será acionado. Esse <i>template</i>, inicialmente, era chamado de placar falso.</p>	<p>= FG_M) e vice versa</p> <ul style="list-style-type: none"> <li>• Onde <math>IND\_F\_VIS = (4 * FG\_M + 2 * FT\_M + 2 * FD\_M + FF\_M + FB\_M) / 10</math></li> <li>• Onde <math>IND\_F\_VIS = (4 * FG\_V + 2 * FT\_V + 2 * FD\_V + FF\_V + FB\_V) / 10</math></li> <li>• Onde  FG_M = finalização gol feita pela equipe mandante  FT_M = finalização na trave feita pela equipe mandante  FD_M = finalização defendida por goleiro feita pela equipe mandante  FF_M = finalização pra fora feita pela equipe mandante  FB_M = finalização bloqueada por algum jogador feita pela equipe mandante</li> <li>• Onde  FG_V = finalização gol feita pela equipe visitante  FT_V = finalização na trave feita pela equipe visitante  FD_V = finalização defendida por goleiro feita pela equipe visitante  FF_V = finalização pra fora feita pela equipe visitante  FB_V = finalização bloqueada por algum jogador feita pela equipe visitante</li> </ul>
---	---

**Tabela 6 - Estereótipo Placar Duvidoso**

As principais alterações feitas nas regras de detecção deste estereótipo foram para ajustar os índices de finalização (IND\_F\_MAN, IND\_F\_VIS), a diferença percentual entre eles, e os valores mínimos para as defesas difíceis (DD\_M, DD\_V).

A primeira versão dos índices de finalização não considerava as finalizações que resultaram em gol. As finalizações gol eram consideradas dentro da regra porém fora dos índices de finalização. Nesta fase a regra detectou jogos, indevidamente, quando o time vencedor fazia muitos gols. Por exemplo, em um jogo em

que o time mandante ganhou por 6 a 0 do time visitante. Suponha-se que o time visitante teve um total de finalizações não resultantes em gol trinta por cento maior que o total de finalizações não resultantes em gol do time mandante. Além disso, o time mandante teve trinta por cento a mais de defesas difíceis do que o time visitante. Nesse cenário a regra detectaria o estereótipo de Placar falso mesmo tendo sido uma goleada. No entanto, ao incluir as finalizações convertidas em gol no índice de finalizações, o time mandante passaria a ter um índice de finalizações maior devido ao grande número de gols e com isso o sistema não detectaria o estereótipo de Placar falso.

A segunda alteração na regra, foi para garantir que a diferença entre os índices de finalização dos dois times seja representativa. Num primeiro momento, a regra verificava apenas se um time tinha esse índice maior do que o outro time. Com isso, uma pequena diferença no índice de finalizações, somada aos outros critérios da regra, detectava o estereótipo Placar Duvidoso. Comparando com a percepção de especialistas foi constatado que a sugestão do estereótipo era indevida.

A última alteração também foi para garantir representatividade na diferença entre os números de defesas difíceis de cada time. A diferença era detectada quando acontecia poucas ocorrências desse *scout* no jogo. Por exemplo, uma defesa difícil é maior que cinquenta por cento de zero defesas difíceis. No entanto, essa diferença não é representativa, então, foi colocado o mínimo de duas defesas difíceis para o time com o menor número de defesas difíceis.

Vejam a seguir quais estereótipos de personagem que foram implementados no protótipo.

Estereótipo Melhor performance do jogo	
Descrição	Regras
Melhor performance em um jogo é um <i>template</i> que é acionado para atuações em que um jogador é o que mais desempenhou um determinado <i>scout</i> em um jogo, e esse desempenho foi, significativamente, maior do que a média do campeonato.	<ul style="list-style-type: none"> <li>• Atuações recordes de um determinado <i>scout</i> em um jogo</li> <li>• Performance acima da média. Atuação que se enquadra nas 5% melhores atuações daquele <i>scout</i> na edição do campeonato</li> <li>• A lista completa de <i>scouts</i> usados nos estereótipos de jogo é:</li> </ul>

	finalizações, roubadas de bola, defesas difíceis, assistências, faltas recebidas, faltas cometidas e gols.
--	--

Tabela 7 - Estereótipo Melhor Performande do Jogo

### 4.3.2 Formação dos esquemas narrativos

Esta seção mostra os esquemas narrativos criados para representar os estereótipos apresentados na seção anterior. Esquema narrativo, no contexto deste trabalho, é o conjunto de informações usadas para apresentar algum estereótipo. Aqui chamado de *template* narrativo, pode ser usado na construção de matérias acerca de um jogo de futebol.

Existem vários *templates* narrativos que podem ser sugeridos pelo sistema de acordo com os acontecimentos de um jogo.

Veremos a seguir, para cada estereótipo implementado no protótipo, uma visualização de seus respectivos *templates* narrativos e uma descrição dos mesmos. Serão apresentados, primeiramente, os *templates* de estereótipos de jogo e em seguida os *templates* de estereótipos de personagem.

1. O *template* Jogo movimentado é composto por três itens de informação. O primeiro é a indicação do estereótipo. O segundo é a explicação do estereótipo. O terceiro item deste *template* é chamado aqui de comparativo entre equipes.

<<indicação do estereótipo>>
<<explicação do estereótipo>>
<<comparativo entre equipes>>

Tabela 8 - Template narrativo Jogo Movimentado / Jogo Morno

O sistema usará como valor padrão para o item *indicação do estereótipo* o próprio nome do estereótipo, por exemplo, Jogo Movimentado.

O item de informação *explicação do estereótipo* é um conjunto de informações relevantes para a compreensão do motivo que fez o jogo se enquadrar nesse estereótipo. No caso do *template* Jogo Movimentado é exibido a quantidade de gols e finalizações do jogo e as médias de gols e finalizações por jogo no campeonato para que o leitor possa fazer uma comparação.

O último item de informação exibe um detalhamento das finalizações por equipe, para que o leitor possa entender a performance em finalizações de cada equipe no jogo. O detalhamento é composto por gols, total de finalizações, finalizações na trave e defesas difíceis.

2. O *template* Jogo morno é composto pelos mesmos itens de informação que compõem o *template* de jogo movimentado. O mesmo formato de *template* atende aos dois estereótipos porque eles são antagônicos. Desta forma, para o estereótipo de Jogo Morno, as finalizações e gols comparados às médias do campeonato servem para ilustrar a baixa incidência de finalizações e gols no jogo.
3. Também composto de três itens de informação, o *template* Goleada tem apenas um item em comum com os *templates* anteriores, a identificação do estereótipo.

<<indicação do estereótipo>>
<<ocorrência histórica >>
<<ocorrências recordes na competição>>

**Tabela 9 - Template narrativo Goleada**

O segundo item de informação retrata parte do aspecto histórico dos jogos entre esses dois times em todas as edições do campeonato, e exibe a data e

o placar do jogo da última goleada entre as duas equipes em edições anteriores da competição.

O terceiro item é uma lista de jogos que representa o ranking das maiores goleadas da edição de campeonato corrente até a data do jogo. Desta forma, o leitor pode comparar a goleada detectada com as maiores goleadas da mesma edição de campeonato.

4. A estrutura do *template* Jogo Truncado é igual à de Jogo Movimentado. A diferença está nos dados que compõem os itens de informação.

<<indicação do estereótipo>>
<<explicação do estereótipo>>
<<comparativo entre equipes>>

**Tabela 10 - Template narrativo Jogo Truncado**

Para o item de informação *explicação do estereótipo* os dados usados são: quantidade de faltas, quantidade de cartões e as médias de faltas e cartões por jogo no campeonato.

No item de informação *comparativo entre equipes* foram usadas as quantidades de faltas, cartões amarelos e cartões vermelhos de cada equipe no jogo.

5. No *template* Placar duvidoso foi reservado um item de informação para uso editorial.

<<indicação do estereótipo>>
<<espaço editorial >>
<<comparativo entre equipes>>

**Tabela 11 - Template narrativo Placar Duvidoso**

Embora exista a possibilidade de edição do conteúdo sugerido em todos os *templates*, para o Placar Duvidoso foi criado um item de informação projetado para ser editado a cada uso do *template*. O item de informação *espaço editorial* exibe a frase “Números que não refletem o placar” apenas como marcação.

No item de informação *comparativo entre equipes* foram usadas informações sobre finalizações que, de certa forma, não são condizentes com o placar. O detalhamento é composto de quantidades de finalizações, finalizações na trave e defesas difíceis.

6. O último *template* de estereótipo de jogo implementado no protótipo foi o *Primeiro no ranking do campeonato*, que pode ser usado para tipos de *scout* diferentes. No protótipo esse *template* foi usado para os *scouts* faltas e passes errados.

<<indicação do estereótipo>>
<<explicação do estereótipo>>
<<ocorrências recordes na competição>>

**Tabela 12 - Template narrativo Primeiro no ranking do campeonato**

Para o item *explicação do estereótipo*, é usado a quantidade de ocorrências de um *scout* que caracterizou o recorde da competição. Assim como a média desse *scout* por partida no campeonato. Dessa forma, o leitor consegue comparar a o jogo recorde com a média da competição.

O terceiro item é uma lista de jogos que representa o ranking dos jogos com maior número de ocorrências do *scout* analisado na edição de campeonato corrente até a data do jogo.

7. Para os estereótipos de personagem foi feito apenas um modelo de *template* que é usado para vários tipos de *scout* diferentes. Assim como em todos os *templates*, o primeiro item de informação é a identificação do estereótipo.

<<indicação do estereótipo>>
<<identificação do personagem>>
<<ocorrências recordes na competição>>

**Tabela 13 - Template narrativo Atleta em Destaque**

O segundo mostra a identificação do atleta que se destacou e a quantidade de ocorrências de um determinado *scout* que o colocou em destaque. Serve para identificar quem e porque se enquadrou no estereótipo.

O terceiro item exhibe o ranking de atuações recordistas de um determinado *scout*, para que o leitor possa comparar a performance detectada com as maiores performances daquele mesmo *scout* na mesma edição de campeonato. Todas as atuações exibidas nesta lista são instâncias do estereótipo de personagem em outros jogos da competição.

No próximo capítulo será mostrado o protótipo feito com base nos estereótipos, seus respectivos templates narrativos e na ontologia de futebol, todos apresentados neste capítulo. Na seção 5.1.2 Exemplos de templates de estereótipos são exibidas algumas figuras com o *templates* sugeridos pelo protótipo com dados reais.

## 5 Estudo de caso

Após a realização de todos os estudos descritos nos capítulos anteriores, foi feito um protótipo para avaliar na prática o objeto investigado.

Um grupo seletivo de profissionais que trabalham na redação da organização onde o experimento foi realizado, formado por redatores e editores, utilizou o protótipo em ambiente real. Alguns profissionais já haviam sido envolvidos nas etapas anteriores desta pesquisa, outros foram envolvidos somente na etapa do estudo de caso com utilização do protótipo. Para estes, foi feita uma apresentação sobre os principais conceitos e objetivos da pesquisa em questão. O conceito de produção de conteúdo para *web* com suporte semântico foi compartilhado pelo grupo.

Nas seções deste capítulo serão descritos o protótipo funcional usado e a experiência de uso do mesmo.

### 5.1 Protótipo funcional

#### 5.1.1 Interface de autoria de matéria

Na organização onde a pesquisa foi realizada já existia um sistema de publicação para *web*, que dentre outras funcionalidades, dispunha de uma interface de autoria de matérias. Uma das ferramentas mais usadas para produção de conteúdo dentro da organização.

Essa interface é um sistema *web* para cadastro de matérias que tem embutido um *plugin* para edição de texto chamado *ckeditor*. Utilizado no campo correspondente ao corpo da matéria, permite ao editor formatar o texto, inserir elementos como fotos e vídeos, criar links para outros conteúdos, dentre outras funcionalidades. Este *plugin* desenvolvido em *Javascript* é extensível e permite o desenvolvimento de funcionalidades customizadas para manipulação do conteúdo que é inserido na área de edição de texto.

Foi desenvolvido como protótipo funcional uma extensão para o editor de texto existente. Esta extensão dá suporte à construção de narrativas acerca de jogos de



futebol. Ela permite ao usuário verificar se o sistema detectou algum estereótipo interessante sobre o jogo. Também disponibiliza *templates HTML* com texto e imagem que caracterizam os estereótipos detectados e podem ser usados dentro dos textos das matérias.

A utilização dos *templates* acontece dentro do cadastro de matéria de forma semi-automática e com decisão editorial. Ou seja, o sistema sugere mas a decisão de usá-los ou não na matéria é do redator. Além disso, uma vez que um *template* foi inserido no texto é possível editar as informações do mesmo se for conveniente. Existe um botão no painel de ferramentas do editor de texto que ao ser clicado é exibida uma tela para detecção de estereótipos. Este botão só deve ser usado para produção de matérias que tratem, especificamente, de jogos de futebol.

Veja na figura Figura 24 como é a interface de autoria de matéria. Todas os elementos da interface exibida, exceto o botão destacado em vermelho, já existiam antes deste trabalho e não serão abordados. Apenas o que o protótipo agregou a ferramenta existente será analisado.

Criar matéria

Listar matérias » Criar outra matéria »

Título da matéria

Subtítulo da matéria

Url (seo):

B I H1 H2 H3 H4 H5 H6 Link Imagem Vídeo YouTube Embed Social Media Undo Redo Bold Italic Underline Strikethrough Bulleted List Numbered List Indent Outdent Align Left Center Right Justify Full Screen Help

Informações da matéria

Editoria

Assinatura

Fonte da notícia

Origem da notícia

Anotação semântica

Esportes (Esporte, Categoria e Modalidade)

Pessoas (Atleta, Árbitro e Técnico)

Organizações (Clube, Agremiação, Time e Seleção)

Eventos (Edição de Campeonato, Jogo e Prova)

Cita o Canal ou Programa de TV

Trata do Assunto

Agendar publicação

salvar preview publicar

**Figura 24 - Interface de autoria de matérias.**

Ao clicar no botão indicado a tela de detecção de estereótipos é exibida por cima da tela de cadastro de matérias. No primeiro momento é apresentado apenas um campo para que o usuário informe a data do jogo sobre o qual a matéria será escrita. Após o preenchimento deste campo com uma data específica será exibida uma lista de jogos que acontecem na data indicada. Veja na Figura 25 - Interface para seleção de jogo para crônica. uma imagem da tela já com a lista de jogos exibida.

The screenshot shows a web interface titled "Informações de confronto" (Match Information) on the globo.com website. The interface is for selecting a match for a chronicle. It features a date field labeled "Data do confronto" with the value "2011-11-20". Below this is a section labeled "Confronto" with a dropdown menu that says "Selecione o confronto". The dropdown is open, displaying a list of matches with their scores: "Corinthians 2 x 1 Atlético-MG" (highlighted), "Botafoogo 1 x 2 Internacional", "Cruzeiro 1 x 1 Atlético-PR", "Figueirense 0 x 4 Fluminense", "Coritiba 1 x 0 Santos", "Atlético-GO 0 x 0 Flamengo", and "Bahia 0 x 2 Palmeiras". At the bottom of the form, there are two buttons: "usar" (use) and "cancelar" (cancel).

Figura 25 - Interface para seleção de jogo para crônica.

Quando o usuário seleciona o jogo desejado, o sistema inicia a detecção de estereótipos. As regras de detecção usadas foram descritas na seção 4.3.1 Regras de detecção de estereótipos. Caso o sistema identifique estereótipos para o jogo indicado, seus respectivos *templates HTML* são exibidos na tela para que o usuário possa selecionar aqueles que julgar adequados para a matéria. Veja na Figura 26 - Interface de sugestão de estereótipos, uma imagem da tela de detecção de estereótipos que exibe alguns *templates HTML* sugeridos para o jogo indicado.

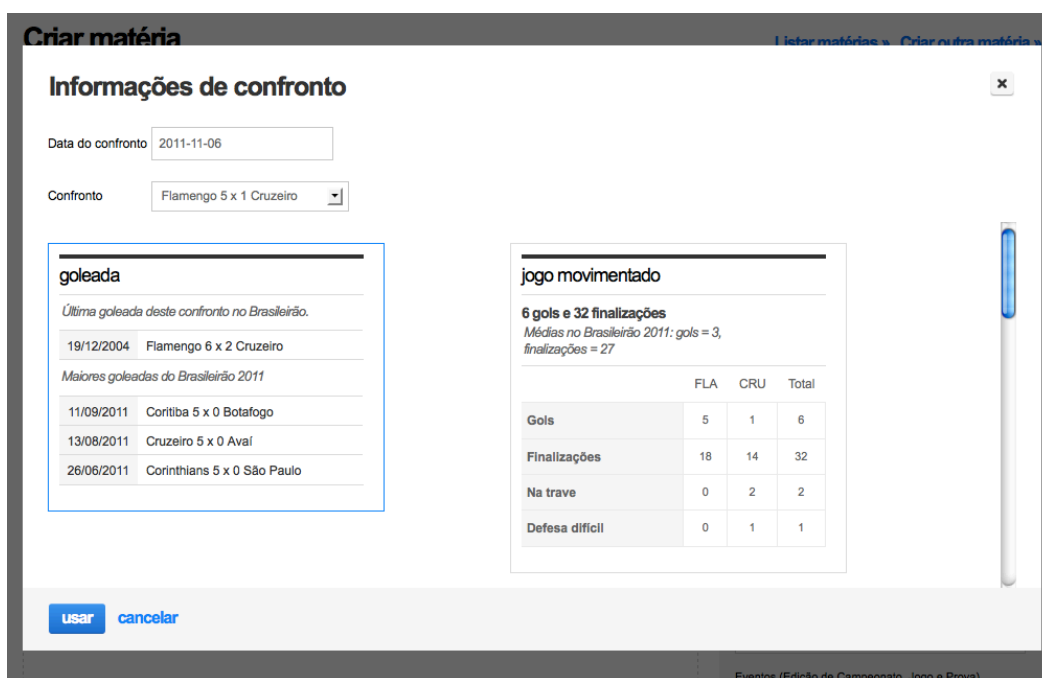


Figura 26 - Interface de sugestão de estereótipos.

Cada *template* sugerido é exibido exatamente como será exibido na matéria final que será publicada na *web* para os internautas. A única diferença é que na tela de detecção de estereótipos os *templates* aparecem com uma linha de contorno cinza para delimitar a área de cada *template*. Essa linha de contorno não é exibida na matéria final.

Quando o usuário decide usar um determinado *template*, ele deve clicar em cima do mesmo para selecioná-lo. Neste momento, a linha de contorno do *template* muda para a cor azul para colocar em destaque o que foi selecionado. Veja na Figura 26 a exibição do *template* de goleada após ter sido selecionado pelo usuário. Após selecionar um ou mais *templates* o usuário deve clicar no botão “usar” para que o conteúdo selecionado seja inserido no corpo da matéria, na posição em que o cursor estiver localizado dentro do editor de texto. Veja na Figura 27 como fica o conteúdo do editor de texto após a inclusão de um *template* de estereótipo.

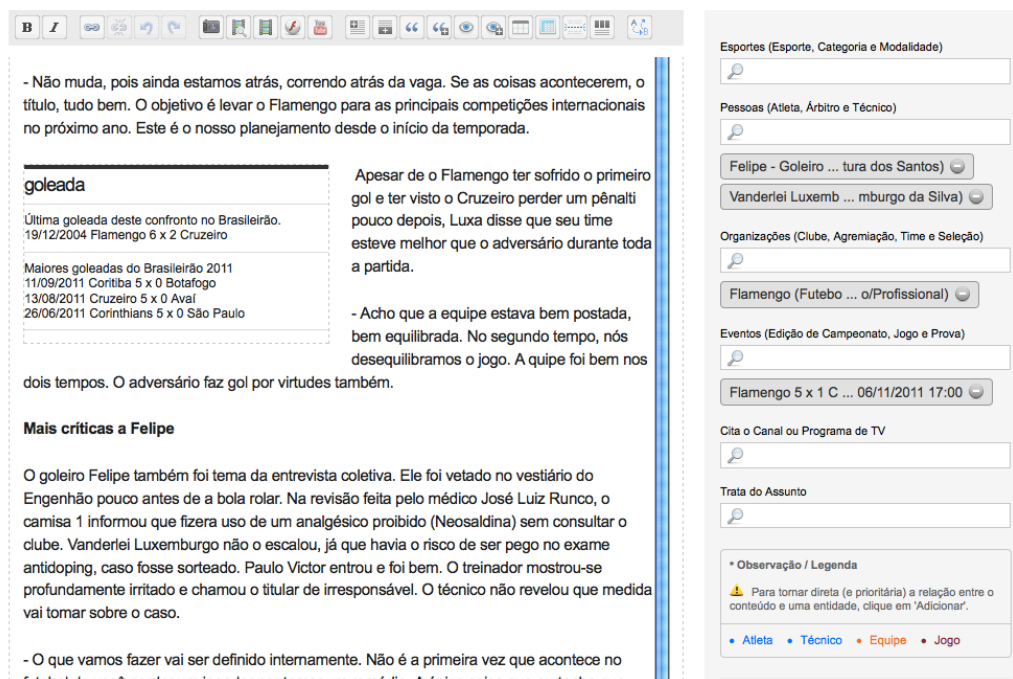


Figura 27 - Interface de autoria de matéria com template de estereótipo selecionado.

O usuário que está criando a matéria ainda tem a possibilidade de editar o conteúdo exibido no *template*. Por exemplo, é possível editar o *label* “goleada” por “mais uma goleada” ou qualquer outra informação desejada. Isso vale para todos os elementos do *template*.

Embora seja possível editar o conteúdo sugerido, foi feito um levantamento de quais informações seriam interessantes para cada estereótipo. A ideia é minimizar o trabalho durante a construção das narrativas. Na seção a seguir apresentamos uma ilustração visual e uma explicação textual de todos os *templates* implementados no protótipo.

### 5.1.2 Exemplos de templates de estereótipos

Veremos nessa seção algumas figuras que exibem a apresentação final dos *templates* de estereótipos instanciados com dados reais. Além da visualização de algumas matérias produzidas com os *templates*.

## 1) Jogo Movimentado

jogo movimentado			
<b>9 gols e 29 finalizações</b>			
<i>Médias no Brasileirão 2011: gols = 2, finalizações = 26</i>			
	SAN	FLA	Total
<b>Gols</b>	4	5	9
<b>Finalizações</b>	13	16	29
<b>Na trave</b>	0	0	0
<b>Defesa difícil</b>	3	2	5

Figura 28 - Template narrativo para estereótipo de jogo movimentado.

A Figura 30 exibe a o template Jogo Movimentado instanciado para o jogo entre Santos e Flamengo, pela décima segunda rodada do campeonato brasileiro de 2011. Vale notar, que além de ter sido movimentado, o jogo foi bem disputado, com números de gols, finalizações, finalizações na trave e defesas difíceis parecidos entre os dois times.

## 2) Jogo Morno

jogo morno			
<b>0 gols e 12 finalizações</b>			
<i>Médias no Brasileirão 2011: gols = 3, finalizações = 27</i>			
	COR	PAL	Total
<b>Gols</b>	0	0	0
<b>Finalizações</b>	5	7	12
<b>Na trave</b>	0	1	1
<b>Defesa difícil</b>	0	0	0

Figura 29 - Template narrativo para estereótipo de jogo morno.

A Figura 29 mostra o *template* de Jogo Morno (igual ao de Jogo Movimentado) porém com quantidades de finalizações e gols reduzidas (característica deste estereótipo), bem menores do que as médias do campeonato. Instanciado para o jogo entre Corinthians e Palmeiras pela trigésima oitava rodada do campeonato brasileiro de 2011.

### 3) Goleada

goleada	
<i>Última goleada deste confronto no Brasileirão.</i>	
19/12/2004	Flamengo 6 x 2 Cruzeiro
<i>Maiores goleadas do Brasileirão 2011</i>	
11/09/2011	Coritiba 5 x 0 Botafogo
13/08/2011	Cruzeiro 5 x 0 Avaí
26/06/2011	Corinthians 5 x 0 São Paulo

Figura 30 - Template narrativo para estereótipo de goleada.

A Figura 30 mostra o *template* Goleada após ser instanciado com o jogo entre Flamengo e Cruzeiro, pela trigésima terceira rodada do campeonato brasileiro de 2011.

### 4) Placar Duvidoso

placar duvidoso		
<i>Números que não refletem o placar</i>		
	BOT	FLU
Finalizações	22	6
Na trave	0	0
Defesa difícil	1	4

Figura 31 - Template narrativo para estereótipo de placar duvidoso.

A Figura 31 exibe o *template* Placar Duvidoso instanciado com o jogo entre Botafogo e Fluminense pela trigésima oitava rodada do campeonato brasileiro de 2011. Esse jogo terminou empatado com placar de 1 a 1, no entanto, os números exibidos pelo *template* mostram que o Botafogo foi mais ofensivo.

## 5) Matéria sobre o jogo Flamengo 5 x 1 Cruzeiro pela rodada 33

06/11/2011 20h21 - Atualizado em 06/11/2011 22h32

# Luxa não se empolga com goleada e mantém meta: vaga na Libertadores

Técnico volta a criticar goleiro Felipe e diz que vai avaliar caso internamente

Por Richard Souza  
Rio de Janeiro

[Tweeter](#)

80

[Recomendar](#)

105



Vanderlei Luxemburgo não se deixou levar pela vitória empolgante do Flamengo. Neste domingo, o time goleou o Cruzeiro por 5 a 1 no Engenhão e voltou com força à briga pelo título brasileiro. Com 55 pontos, está a três do líder Corinthians, que foi derrotado pelo lanterna América-MG (2 a 1). O treinador fez questão de frisar que o discurso não vai mudar. A prioridade é, pelo menos por enquanto, a vaga na Libertadores da América.

- Não muda, pois ainda estamos atrás, correndo atrás da vaga. Se as coisas acontecerem, o título, tudo bem. O objetivo é levar o Flamengo para as principais competições internacionais no próximo ano. Este é o nosso planejamento desde o início da temporada.

### goleada

Última goleada deste confronto no Brasileirão.

19/12/2004 Flamengo 6 x 2 Cruzeiro

Maiores goleadas do Brasileirão 2011

11/09/2011 Coritiba 5 x 0 Botafogo

13/08/2011 Cruzeiro 5 x 0 Avaí

26/06/2011 Corinthians 5 x 0 São Paulo

Apesar de o Flamengo ter sofrido o primeiro gol e ter visto o Cruzeiro perder um pênalti pouco depois, Luxa disse que seu time esteve melhor que o adversário durante toda a partida.

- Acho que a equipe estava bem postada, bem equilibrada. No segundo tempo, nós desequilibramos o jogo. A quipe foi bem nos dois tempos. O adversário faz gol por virtudes também.

Figura 32 - Trecho de matéria feita com template Goleada



## 6) Matéria sobre o jogo Botafogo 1 x 1 Fluminense pela rodada 38

04/12/2011 21h03 - Atualizado em 05/12/2011 10h31

### Cortês dispara: 'O time cria, mas não marca, e acaba estourando lá atrás'

Lateral deixa o campo chateado com o empate contra o Fluminense

Por GLOBOESPORTE.COM  
Volta Redonda, RJ

Tweetar

15

Recomendar



O empate em 1 a 1 com o Fluminense, na última rodada do Brasileiro, deixou os jogadores do Botafogo mais abatidos (assista aos melhores momentos do clássico no vídeo ao lado). Com a certeza de que fizeram uma boa partida, os atletas deixaram o campo até mesmo irritados. Ao tentar explicar por que a equipe não conseguiu o resultado positivo, Cortês reclamou da falta de pontaria do setor ofensivo da equipe.

- Foi o que aconteceu no campeonato todo. O time cria, mas não marca, e acaba estourando lá atrás – desabafou o lateral-esquerdo na saída de campo.

Mais comedido, o volante Renato preferiu analisar o empate por outro ângulo. Para o jogador, faltou um pouco de sorte ao time alvinegro durante o duelo contra o rival tricolor neste domingo, em Volta Redonda.

#### o que o placar não diz

*Alvinegros concluem bem mais no jogo*

	BOT	FLU
Finalizações	22	6
Na trave	0	0
Defesa difícil	1	4

- Infelizmente, mais uma vez as oportunidades que nós criamos criou não entraram. Mas fizemos um bom jogo – disse Renato.

Os jogadores do Botafogo saem de férias agora e só retornam no dia 4 de janeiro, quando tem início a primeira etapa da pré-temporada. O time vai passar cerca de uma semana no Rio e depois viaja para a Região dos Lagos para fazer os últimos dias de preparação em Saquarema.

Figura 33 - Trecho de matéria feita com template Placar Duvidoso

Na Figura 33 é possível notar que não foi usado o valor padrão do *template* para o item de informação *indicação de estereótipo*. O redator que criou a matéria optou por trocar o texto “placar duvidoso” por “o que o placar não diz”. Para o item de informação *espaço editorial*, o editor usou o texto “Alvinegros concluem bem mais no jogo”.

## 7) Matéria sobre o jogo Fluminense 5 x 4 Grêmio pela rodada 35

16/11/2011 22h37 - Atualizado em 17/11/2011 00h19


### Autor de quatro dos cinco gols do Flu, Fred destaca atuação de Deco

'É ele quem acelera, cadencia, dita o ritmo da equipe inteira', diz atacante, que também elogia determinação da equipe do início ao fim

Por **GLOBOESPORTE.COM**  
Rio de Janeiro

 **Tweetar**

88

 **Recomendar**

108



Herói da vitória por 5 a 4 sobre o Grêmio, Fred agradeceu a Deco pelo passe longo no segundo gol tricolor, quando os visitantes venciam por 2 a 1. Para o autor de quatro dos cinco gols tricolores na noite desta quarta, o meia também merece grande parte dos méritos pela vitória que manteve o Fluminense em terceiro lugar (veja o vídeo com os gols).

- É ele quem acelera, cadencia, dita o ritmo da equipe inteira. Tem feito a diferença.

Todo jogo ele me deixa na cara do gol. É parceiro - comentou Fred.

Abraçado ao atacante na descida para o vestiário, Deco retribuiu o carinho.

#### o artilheiro do jogo



**Fred**  
4 gols

Recordes de gols em um jogo neste Brasileiro

Fred	4 gols	FLU 5 x 4 GRE
Neymar	4 gols	SAN 4 x 1 CAP
André Lima	3 gols	GRE 4 x 0 CAP
Thiago Neves	3 gols	FLA 5 x 1 CRU
Anderson Aquino	3 gols	CFC 5 x 1 VAS
Liedson	3 gols	COR 5 x 0 SPO
Ronaldinho Gaúcho	3 gols	SAN 4 x 5 FLA
Diego Souza	3 gols	CRU 0 x 3 VAS
Fred	3 gols	FLU 3 x 1 CFC
Leandro Damilão	3 gols	PAL 0 x 3 INT

- Ele se movimenta, também facilita.

Além de enaltecer Deco, Fred também fez elogios ao grupo e à torcida pelo apoio no Engenhão. O atacante se surpreendeu com a postura dos tricolores na arquibancada. Na avaliação do camisa 9, o clima era para ser negativo depois da derrota por 2 a 1 para o América-MG no último sábado.

- Depois de uma derrota como aquela, era difícil. E o torcedor jogou junto na dificuldade. Levamos um gol cedo, mas continuaram acreditando. Tem de ser assim até o final. Vamos ver se a gente consegue esse título. Essa vitória nos deixa vivos. Todo o time se empenhou muito - disse o atacante.

Fred lamentou apenas a postura dos gremistas, que ameaçaram deixar o campo depois do quinto gol tricolor. Logo após a jogada, Brandão alegou ter sido derrubado com falta por Leandro Euzébio, e gritou muito com o juiz. O

atacante gremista recebeu vermelho e alguns jogadores, indignados, caminharam em direção à lateral para boicotar a partida.

Figura 34 - Trecho de matéria feita com template Artilheiro

## 5.2 A experiência de uso

Durante as últimas seis rodadas do campeonato brasileiro de 2011 o protótipo descrito na seção anterior foi utilizado por cinco jornalistas na construção de matérias sobre os jogos. Embora o grupo seja composto por cinco jornalistas, existe uma escala de plantão que, na prática, permitiu a participação de três jornalistas por rodada.

Como todos os estereótipos disponíveis no protótipo são detectados por regras baseadas em estatísticas de *scouts* (eventos do jogo), e atualmente esses dados são obtidos somente após o término dos jogos, o uso do protótipo foi feito, exclusivamente, para criação de matérias de pós jogo. Nem todos os jogos tiveram suas coberturas jornalísticas feitas pelo grupo que participou do experimento.

O protótipo usado é uma pequena implementação que viabiliza um processo diferente de produção de conteúdo, no qual o jornalista continua a escrever matérias mas ao mesmo tempo interage de forma diferente com o sistema, através da seleção e da organização de conteúdos para a narrativa que está sendo criada por ele.

Essa experiência de uso deve servir para avaliar o fluxo de trabalho proposto e também o protótipo desenvolvido. Em relação ao fluxo, os principais aspectos analisados são:

- 1) A produção de conteúdo para web, de forma semi-automática com decisão editorial é inovadora em relação ao processo usado hoje;
- 2) Esse tipo de produção pode enriquecer o conteúdo de tal forma a superar a qualidade dos conteúdos/produtos atuais;
- 3) O fluxo proposto interfere na produção de conteúdos específicos aumentando a produtividade dos usuários (jornalistas).
- 4) A utilização de *templates* narrativos interfere, de alguma forma, no aumento da “homogeneidade” da qualidade nas matérias, estabelecendo um “estilo editorial”.

Em relação ao protótipo os principais aspectos analisados são:

- 1) A quantidade de estereótipos que o sistema detecta para os jogos;

- 2) Os estereótipos detectados pelo sistema são relevantes e adequadas;
- 3) O conteúdo exibido no *template HTML* dos estereótipos ajuda na criação da narrativa.

As próximas duas seções trazem informações numéricas sobre o uso do protótipo e informações obtidas com uma pesquisa qualitativa feita com os usuários que participaram do experimento.

### 5.2.1 Indicadores sobre a utilização do protótipo

Durante seis rodadas do campeonato brasileiro, foram produzidas duzentas e vinte e quatro matérias, das quais, trinta e três usaram os *templates* do protótipo. Cabe observar que dessas matérias, apenas quarenta jogos foram cobertos por jornalistas envolvidos no experimento. Essas matérias tratam de quarenta jogos diferentes. Veja na figura abaixo um gráfico da quantidade de matérias criadas por rodada, e a quantas destas usaram *templates* do protótipo.

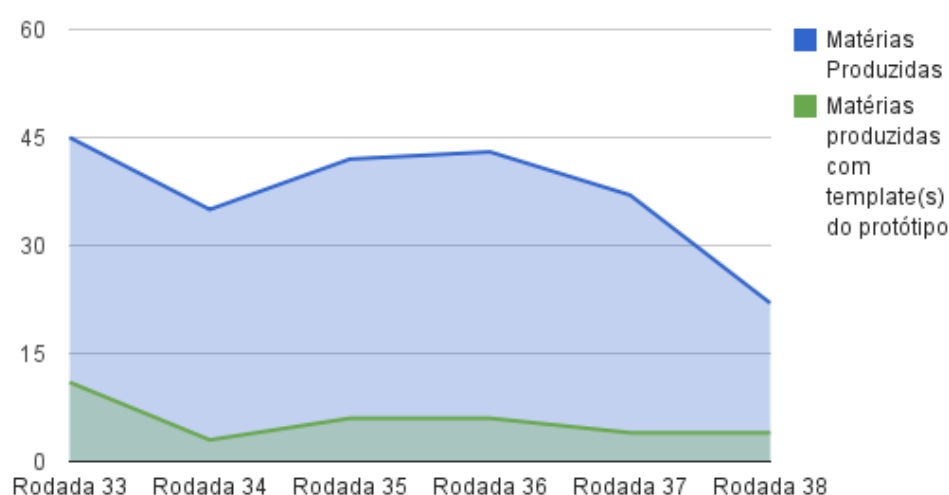


Figura 35 - Gráfico comparativo entre quantidade de matérias produzidas e quantidade de matérias produzidas com templates narrativos do protótipo.

A rodada trinta e três, a primeira em que o protótipo foi usado, foi a rodada com o maior número de matérias feitas com uso de *templates* do protótipo. Ao

todo foram onze matérias, de um total de quarenta e cinco matérias publicadas para os jogos. Isto representa que, aproximadamente, vinte e cinco por cento das matérias produzidas para os jogos daquela rodada usaram os *templates HTML* sugeridos pelo protótipo.

Na rodada seguinte, o uso foi menos intenso, foram publicadas somente três matérias com *templates HTML*, de um total de trinta e cinco que foram publicadas para os jogos desta rodada. Apenas oito e meio por cento do total.

Considerando todas as seis rodadas, pode-se dizer que, aproximadamente, quinze por cento das matérias de pós jogo produzidas neste período usaram os *templates HTML* sugeridos pelo protótipo.

O número de matérias por rodada varia, de acordo, com as repercussões dos jogos, que por sua vez depende dos acontecimentos durante as partidas. Partidas com muitos gols envolvendo algum time grande ou partidas com boas atuações de atletas famosos, geram mais matérias.

Veja na figura a seguir um gráfico que mostra a quantidade de jogos por rodada que a equipe de jornalistas envolvida fez cobertura. E quantos desses jogos tiveram pelo menos uma matéria feita com o uso do protótipo.

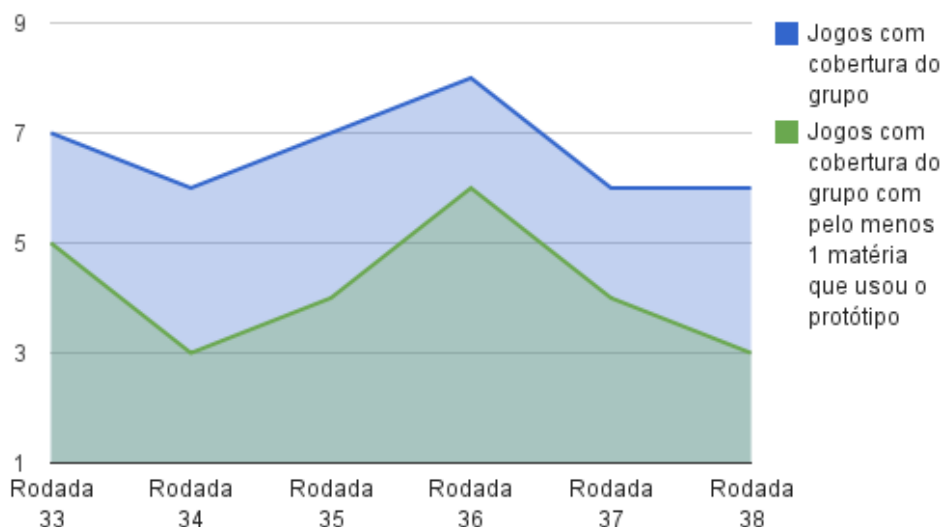


Figura 36 - Gráfico comparativo entre quantidade de jogos e a quantidade de jogos para os quais pelo menos uma matéria foi produzida com template narrativo do protótipo.

Ao todo quarenta jogos tiveram suas coberturas jornalísticas feitas pelo grupo. Para vinte e cinco deles, pelo menos uma matéria foi feita com algum *template HTML* do protótipo. Ou seja, em sessenta e dois e meio por cento dos jogos, para os quais o protótipo foi dedicado, foi feita pelo menos uma matéria com o suporte do mesmo.

Na rodada trinta e seis oito jogos foram acompanhados pelo grupo, para seis deles, foi feita pelo menos uma matéria com o uso do protótipo. Ou seja, em sessenta e cinco por cento dos jogos, o protótipo foi útil para pelo menos uma matéria sobre o jogo. Essa foi a rodada que teve o maior índice neste aspecto.

As rodadas trinta e quatro e trinta e oito foram as que tiveram os piores índices neste aspecto, cinquenta por cento. Em ambas rodadas o grupo fez a cobertura de seis jogos, mas em apenas três deles o protótipo foi útil para alguma matéria.

Outra forma de analisar o comportamento do protótipo é verificar dentre as indicações de estereótipos feitas por ele quantas foram efetivamente usadas em alguma matéria. Durante o experimento, o protótipo sugeriu oitenta e seis *templates* de estereótipos diferentes para os quarenta jogos. Desses oitenta e seis *templates* trinta e quatro foram aproveitados. Isso significa que, aproximadamente, quarenta

por cento dos *templates* de estereótipos sugeridos foram aproveitados nas matérias.

Alguns números obtidos na pesquisa dizem respeito à quantidade de sugestões que o protótipo faz por jogo. A média de sugestão de *templates* por jogo foi de aproximadamente dois e meio. O número máximo de indicações de *templates* feitas pelo protótipo em um jogo foi seis. Esta quantidade de indicações ocorreu em dois jogos. Aconteceram quatro jogos, para os quais o sistema não detectou nenhum estereótipo e então nenhum *template* foi sugerido.

A figura a seguir mostra um gráfico com a quantidade de sugestões de cada estereótipo implementado no protótipo e a quantidade de usos que foram feitos de cada um.

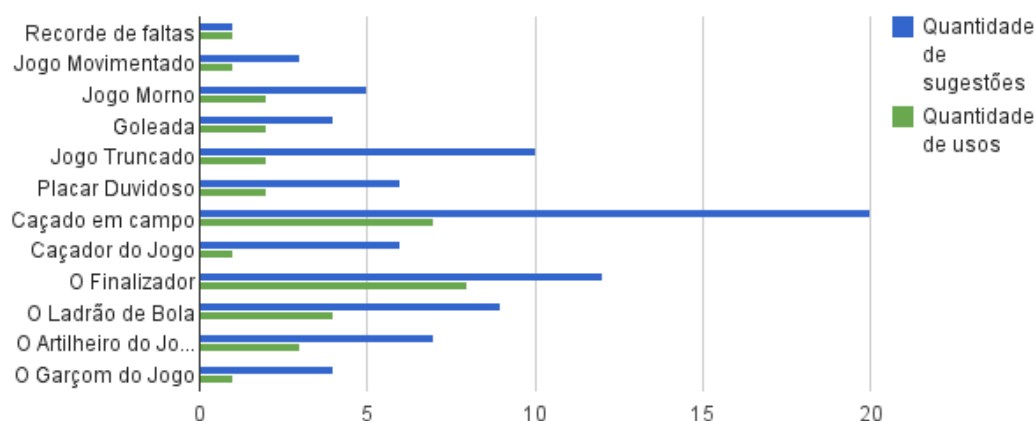


Figura 37 – Gráfico com as quantidades de uso dos templates narrativos do protótipo.

Os seis primeiros estereótipos da lista são os chamados estereótipos de jogo. Dentre eles o que teve o maior índice de aceitação foi o “Goleada”, que das quatro vezes que foi sugerido, em duas foi utilizado.

Os outros seis são estereótipos de personagem. Deste grupo, o que teve o maior índice de aceitação foi o “O Finalizador”, que de doze indicações foi usado oito vezes.

O estereótipo que foi detectado mais vezes foi o “Caçado em campo” que foi sugerido vinte vezes. Este estereótipo é detectado quando algum jogador recebeu em um jogo um número de faltas, significativamente, maior do que a média de faltas recebidas por jogo. E além disso o jogador foi o que mais recebeu faltas no jogo.

Já o estereótipo mais usado foi o “O Finalizador”, que é detectado quando um jogador é o que mais finalizou para o gol na partida e além disto essa quantidade de finalizações é, significativamente, maior do que a média de finalizações por partida de um jogador.

O estereótipo “Recorde de passes errados” que é detectado quando um jogo tem o maior número de passes errados no campeonato não foi detectado nenhuma vez. O mesmo aconteceu para o estereótipo “A muralha do jogo”, que é usado para atuações de goleiros que tenham se destacado pela quantidade de defesas. Ambos não aparecem no gráfico acima.

### 5.2.2 Pesquisa de opinião sobre o uso do protótipo

Foi feita uma pesquisa de opinião sobre o uso do protótipo com os jornalistas que participaram do experimento. Antes da pesquisa, propriamente dita, a ideia de produção de conteúdo para web com suporte semântico foi, novamente, discutida pelo grupo. Em seguida, foi criado um questionário na *web*, que foi respondido pelos jornalistas, com questões sobre os conceitos que foram testados e sobre o protótipo em si.

As questões da pesquisa tem como objetivo capturar o quanto o modelo proposto é aderente às necessidades e possibilidades vislumbradas pelos jornalistas. E o quanto o protótipo é aderente ao modelo proposto. Com seis opções de respostas, sendo que *A*) significa totalmente aderente e *F*) significa não aderente. A Tabela 14 apresenta o resultado da pesquisa. Cada linha da tabela representa uma pergunta e as colunas representam as opções de resposta. Nas células são exibidos os totais de respostas dos jornalistas para cada pergunta.



Questões da pesquisa de opinião	A	B	C	D	E	F
A ideia de produção de conteúdo para web, de forma semi-automática com decisão editorial é inovadora em relação ao processo usado hoje.	4	0	0	0	0	0
A ideia de produção de conteúdo para web, de forma semi-automática com decisão editorial pode enriquecer o conteúdo de tal forma a superar a qualidade dos conteúdos/produtos atuais	4	0	0	0	0	0
O protótipo usado neste experimento é aderente a ideia de produção de conteúdo para web, de forma semi-automática com decisão editorial.	3	1	0	0	0	0
O protótipo usado neste experimento interfere na produção de conteúdos específicos aumentando a produtividade dos usuários (jornalistas).	1	3	0	0	0	0
O protótipo usado neste experimento ajuda a criar ou tentar expressar novas ideias através de novos conteúdos.	2	1	1	0	0	0
O protótipo usado neste experimento sugere alguns <i>templates</i> com narrativas padronizadas de acordo com os acontecimentos dos jogos. As sugestões feitas pelo sistema são relevantes e adequadas.	1	1	2	0	0	0
O protótipo usado neste experimento teve duas utilizações distintas. Uma delas foi o enriquecimento de matérias já existentes. Para este tipo de utilização o protótipo mostrou eficiência.	3	1	0	0	0	0
O protótipo usado neste experimento teve duas utilizações distintas. Uma delas foi a geração de pauta para assuntos detectados pelo sistema que seriam despercebidos na ausência do mesmo. Para este tipo de utilização o protótipo mostrou eficiência.	0	4	0	0	0	0
É desejável que o <i>plugin</i> utilizado seja incorporado à sua ferramenta de trabalho.	4	0	0	0	0	0

Tabela 14 - Resultado da pesquisa de opinião feitas com os jornalistas envolvidos

## 6 Conclusões e próximos passos

Este capítulo é dividido em duas seções. A primeira descreve as principais conclusões sobre o trabalho realizado. Na segunda seção é mostrado um conjunto de oportunidades de trabalhos relacionados.

### 6.1 Conclusões

O objetivo desta pesquisa foi verificar se um modelo semântico de publicação de conteúdo para *web* pode, de alguma forma, agregar valor ao modelo usado atualmente na organização onde a pesquisa foi realizada.

As observações feitas durante o experimento foram analisadas sob dois aspectos. O primeiro diz respeito ao fluxo de trabalho proposto, parcialmente, experimentado através do uso do protótipo. O segundo aspecto está relacionado ao protótipo em si, em outras palavras, está relacionado com o conteúdo (ou forma de conteúdo) sugerido pelo protótipo durante o experimento.

As avaliações foram feitas com base nos números de utilização do protótipo e na pesquisa de opinião feita com os usuários envolvidos.

O processo de construção de matérias com o uso do protótipo foi considerado inovador em relação ao processo usado hoje. Do ponto de vista dos recursos usados durante a construção de narrativas, pode-se dizer que pela primeira vez o grupo usou uma ferramenta capaz de influenciar tanto na forma do conteúdo quanto no conteúdo em si.

O protótipo foi usado de duas formas diferentes, para enriquecer matérias existentes e para criar matérias que só foram criadas devido à detecção de estereótipos feitas pelo sistema. Por exemplo, o *template* Goleada foi usado em matérias que seriam escritas independentemente do uso do protótipo. Não é necessário que o sistema detecte esse estereótipo para que o jornalista perceba que o jogo foi uma goleada. No entanto, o *template* relacionado a este estereótipo traz informações interessantes para a narrativa de uma goleada, tais como a última goleada entre os times que jogaram e o ranking das maiores goleadas do campeonato. Estas infor-

mações ajudaram a enriquecer a narrativa, e se fosse necessário pesquisá-las fora do sistema, o jornalista possivelmente não as teria usado. Ou seja, a matéria foi enriquecida com o uso do *template* narrativo. Também aconteceu de uma determinada matéria só ser escrita devido à indicação de *template* feita pelo sistema. Um exemplo desse caso é uma matéria sobre o jogo com o maior número de faltas do campeonato. Esta matéria só foi feita porque o estereótipo “Jogo recorde no campeonato” foi detectado. Desta forma, a sugestão do *template* serviu como indicação de pauta e também como um recurso para enriquecer a narrativa que foi criada.

Além disso, o sistema apresenta informações diferentes para cada jogo, mostrando apenas o que foi destaque, poupando trabalho dos jornalistas na hora de pesquisar números interessantes do jogos.

A funcionalidade que permite que usuário edite as informações contidas nos *templates* foi usada de forma eficiente em várias situações. Isto comprovou, de certa forma, a aplicabilidade do conceito de produção de conteúdo de forma semi-automática com decisão editorial. A Figura 33 mostra uma matéria sobre um jogo para o qual essa funcionalidade foi usada.

Foi verificado também que o mesmo *template*, pode ser usado para descrever situações diferentes. Por exemplo, o *template* “finalizador” é detectado através de uma regra que verifica a quantidade de finalizações para o gol feitas por um atleta em um jogo. Sempre que a quantidade for alta, esse *template* é sugerido. No entanto, as finalizações para o gol são de tipos diferentes: finalização convertida em gol, finalização na trave, finalização pra fora, finalização defendida pelo goleiro e finalização bloqueada por algum jogador. Em uma determinada situação o *template* foi usado para enaltecer a quantidade de finalizações, onde o finalizador da partida fez dois gols e uma finalização na trave. Em outra situação o uso teve como objetivo evidenciar a má pontaria de um atleta porque ele foi o finalizador da partida mas não levou perigo ao gol dos adversários. Todas as finalizações foram pra fora ou bloqueadas por algum jogador.

O experimento revelou que a quantidade de *templates* narrativos implementadas no protótipo foi insuficiente. Os números exibidos na seção 5.2.1 Indicadores sobre a utilização do protótipo mostram que em aproximadamente sessenta e três

por cento dos jogos que poderiam ter suas narrativas feitas com o uso do protótipo, isto de fato aconteceu. Ou seja, para mais da metade dos “jogos potenciais” o protótipo foi usado em pelo menos uma matéria. Ao analisar as matérias sobre os jogos foi possível perceber que se o protótipo detectasse novos tipos de estereótipo este percentual poderia aumentar bastante, tais como, estereótipos relacionados a sequência de resultados ou performances e a mudanças de posição na tabela de classificação, que são comumente destacados nas narrativas. Alguns exemplos: time que passou para faixa de rebaixamento do campeonato, time que está há  $n$  jogos sem perder, jogador que fez gol nos últimos  $x$  jogos.

A pesquisa de opinião feita com os jornalistas envolvidos revelou um alto índice de satisfação. Todos declararam interesse em ter uma ferramenta com as características do protótipo definitivamente incorporada à ferramenta usada atualmente.

## 6.2 Próximos passos

É possível identificar diversas oportunidades de trabalhos relacionados. Essas oportunidades serão organizadas em duas seções para facilitar o entendimento: Oportunidades relacionadas ao modelo de publicação; Oportunidades relacionadas aos esquemas narrativos.

### 6.2.1 Oportunidades relacionadas ao modelo de publicação

Boa parte do esquema de publicação proposto na Figura 10 não foi desenvolvida no protótipo apresentado neste trabalho. Desta forma, ficaram algumas oportunidades para trabalhos futuros:

**Linked Data** – o modelo propõe que o sistema de publicação seja integrado a *LOD (Linked Open data)* para que o produtor de conteúdo possa, de alguma forma, dispor de informação externas a organização que estejam disponíveis na *web*. No protótipo, os *templates* narrativos e seus respectivos estereótipos foram feitos com base, exclusivamente, em dados disponíveis no banco de dados da organização.

**Base semântica de esquemas narrativos** – No modelo proposto, os esquemas narrativos, que no protótipo foram implementados de forma simplificada através dos *templates* narrativos, deveriam ser editados ou criados pelo editor de esquema

narrativos. A ideia é criar um modelo através do qual os editores possam definir como devem ser os esquemas narrativos desejados, e quais devem ser as regras de detecção de estereótipos. Inclusive, permitindo que essas regras e esquemas narrativos possam ser configurados por usuário do sistema.

**Interface de autoria de matéria** – Embora a interface de autoria tenha sido customizada para o experimento, isto foi feito de forma simplificada, viabilizando apenas a seleção de *templates* narrativos sugeridos semanticamente. No entanto, o modelo proposto, prevê que a matéria como um todo seja produzida com suporte semântico. Ou seja, todos os elementos da narrativa podem se basear em sugestões do sistema. Textos, fotos, vídeos, notícias relacionadas, *templates* narrativos e infográficos são elementos da narrativa que podem ser sugeridos pelo sistema a partir de uma análise semântica.

### 6.2.2 Oportunidades relacionadas aos esquemas narrativos

Durante as etapas iniciais deste trabalho foram levantados diversos estereótipos de jogos e de personagens que podem ser visualizados na Figura 23 - Matriz de estereótipos. Porém nem todos foram implementados, ou foram implementados de forma simplificada. Desta forma, algumas oportunidades ficaram em aberto:

**Novos estereótipos** - De acordo com os jornalistas que participaram do experimento alguns estereótipos baseados em sequência de eventos são muito importantes. Sequências de resultados, quebras de jejum (quando um time está há muito tempo sem ganhar), entrada e saída de faixas de classificação são exemplos de estereótipos baseados em sequência de eventos que podem ser modelados e implementados em trabalhos futuros.

**Interação entre estereótipos** – Na implementação do protótipo a detecção de um estereótipo sempre gera a indicação de um mesmo esquema narrativo (através de um *template* narrativo). Por exemplo, sempre que o sistema detecta o estereótipo *Caçado em campo*, o mesmo *template* é sugerido independentemente dos outros estereótipos detectados. No entanto, os jornalistas levantaram a possibilidade de se criar alguns *templates* mais complexos formados pela combinação de estereótipos. Desta forma, caso o estereótipo *Caçado em campo* seja detectado conjuntamente com outros estereótipos, por exemplo, *Jogo Truncado* e *Caçador do*

*jogo*, o sistema poderia sugerir um único esquema narrativo reunindo os estereótipos detectados.

**Serviços de identificação de estereótipos** – A identificação do estereótipo se dá apenas no fluxo de criação de matérias. A possibilidade de se criar um *webservice* para consulta de estereótipos faria com que outras aplicações pudessem utilizar essas informações.

## 7 Referências bibliográficas

[Allemang 2010] ALLEMANG, D., Semantic Web and the Linked Data Enterprise, Linking Enterprise Data, 2010.

[Allen 1983] ALLEN, J.F., Maintaining Knowledge About Temporal Intervals, Communications of the ACM, Vol.26, no. 11, 1983.

[Almeida et al, 2008] ALMEIDA, J.P.A., GUIZZARDI, G. A Semantic Foundation for Role-Related Concepts in Enterprise Modelling. In: THE 12th IEEE INTERNATIONAL EDOC CONFERENCE - EDOC 2008, München, Germany, 2008.

[Berners-Lee et al, 2001] BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web, Scientific American, 2001.

[Bottazzi et al, 2008] BOTTAZZI, E.; FERRARIO, E., Towards a DOLCE Ontology of Organizations, Journal of Business Process Integration and Management (IJBPM), Inderscience Publisher, 2008.

[Ciarlini et al, 2009] CIARLINI, A.E.M.; BARBOSA, S.D.J.; CASANOVA, M.A.; FURTADO, A.L. Event Relations in Plan-Based Plot Composition. In: Proceedings of SBGAMES 2008, Minas Gerais, Brazil, 2008.

[Conte et al, 1995] CONTE, R., CASTELFRANCHI, C., Cognitive and Social Action, UCL Press, London, 1995.

[Dunn 2010] DUNN, S. From Publisher to Platform - How the guardian used content, search, and open source to build a powerful new business model. Disponível em: <<http://www.slideshare.net/openplatform/from-publisher-to-platform-how-the-guardian-used-content-search-and-open-source-to-build-a-powerful-new-business-model>> Acesso em 25 out. 2010.

[Falbo 2003] FALBO, R.A., ODE: Ontology-based Software Development Environment, IX Congreso Argentino de Ciencias de la Computación, La Plata, Argentina, 2003.

[Falbo et al, 2005] FALBO, R.; BERTOLLO, G., Establishing a Common Vocabulary for Helping Organizations to Understand Software Processes, Proceedings

of the 1st Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE), Enschede, The Netherlands, ISSN 1574-0846, 2005.

[FIFA 2010] FIFA, Laws of the game. Disponível em <[http://www.fifa.com/mm/document/affederation/generic/81/42/36/lawsofthegame\\_2010\\_11\\_e.pdf](http://www.fifa.com/mm/document/affederation/generic/81/42/36/lawsofthegame_2010_11_e.pdf)>. Acesso em 20 out. 2011.

[Guarino et al, 2002] GUARINO, N., WELTY, C. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*, 45 (2): 61-65 (2002).

[Guizzardi et al, 2004] GUIZZARDI, G.; WAGNER, G. On A Unified Foundational Ontology and some Applications of it in Business Modeling, Open INTEROP Workshop on Enterprise Modelling and Ontologies for Interoperability, at the 16th Intl. Conf. on Advances in Information Systems Engineering (CAiSE), Latvia, 2004.

[Guizzardi et al, 2005] GUIZZARDI, G., WAGNER, G. Some Applications of a Unified Foundational Ontology in Business Modeling, In *Ontologies and Business Systems Analysis*, M. Rosemann and P. Green (Eds.), IDEA Publisher, 2005.

[Guizzardi 2005] GUIZZARDI, G. *Ontological Foundations for Structural Conceptual Models*, Telematica Instituut Fundamental Research Series no. 15, The Netherlands, 2005, ISBN 90-75176-81-3.

[Guizzardi & Masolo & Borgo, 2006] GUIZZARDI, G.; MASOLO, C.; BORGO, S. In the Defense of a Trope-Based Ontology for Conceptual Modeling: An Example with the Foundations of Attributes, Weak Entities and Datatypes, 25th Intl. Conf. on Conceptual Modeling (ER'2006), Arizona, USA. LNCS Vol. 4215, Springer-Verlag, Berlin, 2006.

[Guizzardi et al, 2008] GUIZZARDI, G., FALBO, R. A., GUIZZARDI, R. S. S. A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software. *IEEE Transactions Latin America*. v. 6, p. 244-251, 2008.

[Guizzardi et al, 2009] GUIZZARDI G, WAGNER G. Using the Unified Foundational Ontology (UFO) as a foundation for general conceptual modeling languages. In: Poli R, editors, *Theory and Application of Ontologies*, vol. 2, Springer-Verlag: Berlin; 2009.



[Heller et al, 2004] HELLER, B., HERRE, H. Ontological Categories in GOL. Axiomathes 14: 71-90, Kluwer Academic Publishers, 2004.

[Lowe 2006] LOWE, E.J., The Four Category Ontology, Oxford University Press, 2006.

[McLuhan 1964] MCLUHAN, M. Understanding Media: the extensions of man. New York: The New American Library. 1964. 318p.

[New York Times 2009] NEW YORK TIMES. First 5,000 Tags Released to the Linked Data Cloud. Disponível em: <<http://open.blogs.nytimes.com/2009/10/29/first-5000-tags-released-to-the-linked-data-cloud/>> Acesso em 10 out. 2010.

[Ó Cruaiaich 2010] O CRUALAOICH, D. Linked Data at the Guardian. Disponível em : < <http://www.guardian.co.uk/open-platform/blog/linked-data-open-platform>> Acesso em 22 out. 2010.

[O' Donovan 2010] O'DONOVAN, J. The World Cup and a call to action around Linked Data. Disponível em: <[http://www.bbc.co.uk/blogs/bbcinternet/2010/07/the\\_world\\_cup\\_and\\_a\\_call\\_to\\_ac.html](http://www.bbc.co.uk/blogs/bbcinternet/2010/07/the_world_cup_and_a_call_to_ac.html)> Acesso em 20 out. 2010.

[Rayfield 2010] RAYFIELD, J. BBC World Cup 2010 dynamic semantic publishing. Disponível em: <[http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc\\_world\\_cup\\_2010\\_dynamic\\_sem.html](http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html)> Acesso em 20 nov. 2010.

[Rosenberg 1977] ROSENBERG, S.T., Frame-based Text Processing. Massachusetts: Massachusetts Institute of Technology, 1977.

[Searle 2000] SEARLE, J., Mind, Language and Society, Basic Books, 2000.

[Troncy 2008] TRONCY, R., Bringing The IPTC News Architecture into the Semantic Web In: 7th INTERNATIONAL SEMANTIC WEB CONFERENCE - ISWC'08, Karlsruhe, Germany, 2008. p. 483-498