

DIKW4IoT: Uma abordagem baseada na hierarquia *DIKW* para a construção de grafos de conhecimento para integração de dados de *IoT*

Francisca Jamires da Costa, Caio Viktor S. Avila, Tulio Vidal Rolim,
Rossana Maria de Castro Andrade, Vânia Maria Ponte Vidal

¹ Departamento de Computação, Universidade Federal do Ceará(UFC)
Fortaleza - CE, Brasil

{jamirescostaa, arlaass, tulio.xcrtf,
rossanamcandrade, vania.pvidal}@gmail.com

Abstract. *The integration of sensor data in a knowledge graph enables a knowledge base with explicit and implicit information to be created, so that the data is presented in a structured and linked way. However, knowledge graphs are currently constructed without taking into account the different levels of data and their relationships. According to the DIKW hierarchy, data can be classified as data, information, knowledge and wisdom during its life cycle. This work aims to establish the relationships between the different levels of IoT data according to the DIKW hierarchy and demonstrate how these levels can be present in a knowledge graph. For this, an approach is proposed to create a knowledge graph of IoT data with data representation based on the DIKW hierarchy.*

Resumo. *A integração de dados de sensores em um grafo de conhecimento possibilita que uma base de conhecimento com informações explícitas e implícitas seja criada, de maneira que os dados se apresentem de forma estruturada e vinculada. Entretanto, os grafos de conhecimento atualmente são construídos sem levar em consideração os diferentes níveis de dados e suas relações. Segundo a hierarquia DIKW os dados podem ser classificados como dado, informação, conhecimento e sabedoria durante o seu ciclo de vida. Este trabalho tem como objetivo estabelecer as relações entre os diferentes níveis de dados IoT segundo a hierarquia DIKW e demonstrar como estes níveis podem estar presentes em um grafo de conhecimento. Para isso propõe-se uma abordagem para criação de um grafo de conhecimento de dados IoT com representação dos dados baseado na hierarquia DIKW.*

1. Introdução

Com o aumento de dispositivos conectados a internet muitos dados têm sido produzidos e o conceito de *Web* tem mudado. O crescimento da comunicação sem fio e da computação de borda tem favorecido o surgimento de diferentes ambientes para aplicações e assim o aumento da produção de dados [Gao et al. 2021]. O desenvolvimento de aplicações da Internet das Coisas - *IoT* enfrenta diferentes desafios de interoperabilidade. Em uma única aplicação ou domínio tem-se diferentes desafios de interoperabilidade que vão desde os protocolos utilizados para comunicação dos dados até as tecnologias para o gerenciamento dos dados [Gyrard et al. 2016].

Uma observação de um sensor corresponde a um valor lido e uma anotação temporal. Sem a associação de dados que permitam sua identificação e contexto, estes dados não possuem um valor significativo. Além do que os dados de sensores apresentam heterogeneidade sintática e semântica, isto é, diferem no formato dos dados apresentado e no modelo de informação aplicado.

Para auxiliar na resolução destes problemas, muitos trabalhos utilizam ontologias a fim de definir um vocabulário comum sobre os dados [Le-Phuoc et al. 2016, Alvarez-Coello and Gómez 2021]. Uma ontologia consiste em uma terminologia que descreve conceitos para a formalização de um domínio por meio de classes, propriedades e relacionamentos [Pan et al. 2017].

A integração de dados de sensores por meio de ontologias possibilita a criação de uma base de conhecimento, de maneira que os dados se apresentem de forma estruturada e vinculada. Estes dados ao serem integrados reduzem o esforço do usuário de conhecer os diferentes esquemas de dados e a elaboração de diferentes estruturas para transformação dos dados [Steindl et al. 2019]. O resultado dessa integração de dados pode ser compartilhada e explorada através de um grafo de conhecimento semântico.

Essa integração possibilita que informações mais complexas possam ser obtidas por meio dos dados. Estas informações ao serem analisadas geram um conhecimento em relação ao ambiente ou cenário em que os dados de sensores estão sendo obtidos. Essa relação tênue entre dados, informação e conhecimento foi estabelecida outrora na pirâmide do conhecimento, também denominada como hierarquia *DIKW* (*Data-Information-Knowledge-Wisdom*) [Van Meter 2020].

Todavia, essa relação entre dado, informação, conhecimento e sabedoria não é estabelecida em dados de sensores integrados. Os dados de sensores quando integrados por meio de ontologia podem evoluir e apresentar diferentes níveis de abstrações de dados, desde uma informação até um conhecimento específico obtido a partir de uma tarefa de análise de dados. Integrar o conhecimento gerado pela análise de diferentes serviços e aplicações de dados favorece o acesso ao conhecimento de forma mais ampla e menos burocrática, já que não fica refém à estrutura, sintaxe e semântica de uma aplicação.

Construir grafos de conhecimento de dados *IoT* que levem em consideração os diferentes níveis de dados ainda é um problema em aberto, o que leva a pensar também como o processo de anotação semântica pode ser efetuado e quais vocabulários podem ser usados para contemplar os diferentes níveis de dados. Diante disso, este trabalho tem como objetivo estabelecer as relações entre os diferentes níveis de dados de sensores segundo a hierarquia *DIKW* e demonstrar como estes níveis podem estar presentes em um grafo de conhecimento de dados *IoT*. Para isso propõe-se uma abordagem para criação de um grafo de conhecimento de dados *IoT* com representação dos dados baseado na hierarquia *DIKW*.

O restante deste artigo se apresenta da seguinte forma: a Seção 2 apresenta os diferentes níveis de dados da pirâmide do conhecimento e a relação destes com os dados de sensores. A Seção 3 apresenta a abordagem *DIKW4IoT* para a construção do grafo de conhecimento de dados *IoT*, assim como o *workflow* para guiar a sua construção. A Seção 4 apresenta um estudo de caso que foi desenvolvido seguindo as etapas e tarefas propostas. A Seção 5 apresenta um resumo sobre os trabalhos relacionados. A Seção 6

apresenta a conclusão e os trabalhos futuros.

2. Representação do conhecimento baseado na hierarquia DIKW

A pirâmide do conhecimento também conhecida como hierarquia *DIKW* tem como proposta a caracterização e classificação dos dados em quatro camadas: *data*, *information*, *knowledge* e *wisdom* [Van Meter 2020]. A Figura 1.a apresenta a distribuição destas camadas em uma pirâmide.

A hierarquia *DIKW* estabelece que os *dados* são símbolos gravados e leitura de sinais, em que os símbolos são representados por palavras, números, diagramas e imagens e os sinais incluem sensores e leituras sensoriais de luz, som, etc [Liew 2013]. A *informação*, por sua vez, refere-se a uma mensagem com significado mais relevante e que pode ser utilizada como entrada para decisão e/ou ação, ou seja, a informação são dados que por meio de conexões relacionais e contextuais receberam mais significado [Duan et al. 2017].

O *conhecimento* refere-se a capacidade de agir e a compreensão em relação a algo, isto é, a partir da informação recebida, processada e validada cognitivamente o conhecimento é construído [Duan et al. 2017]. A *sabedoria* está relacionada a possibilidade de avaliar um cenário e discernir entre o certo e o errado, sendo capaz de projetar as consequências de uma ação [Remor et al. 2017].

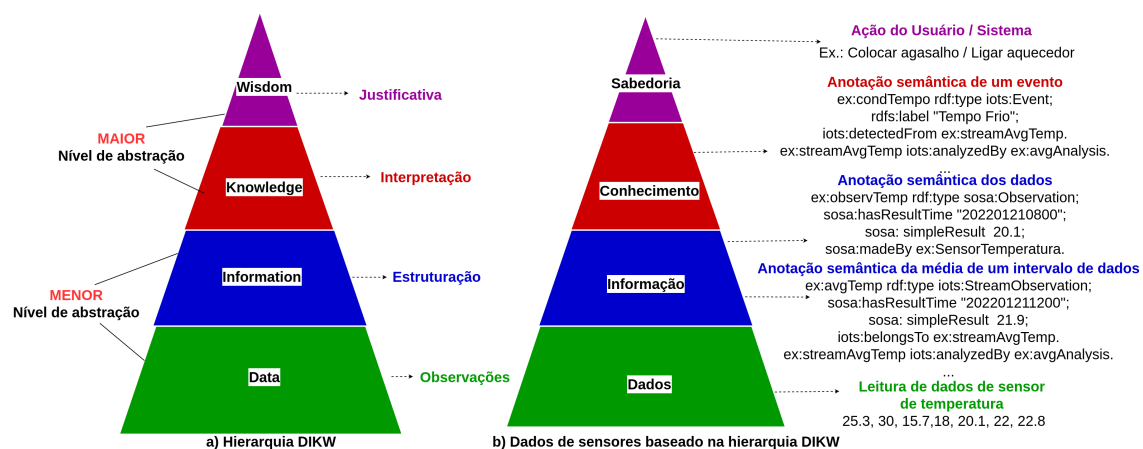


Figura 1. a) Hierarquia *DIKW* - Fonte: Adaptado de [Roberto 2021]
 b) Representação do grafo de Conhecimento de dados *IoT* baseado na hierarquia *DIKW* - Fonte: Elaborado pelo autor

Os dados de sensores também podem ser representados segundo a hierarquia *DIKW*. A Figura 1.b apresenta uma pirâmide de conhecimento de dados de sensores. Uma observação de um sensor faz parte do nível mais baixo da hierarquia *DIKW*, a camada de dados. As leituras de dados de sensores, por si só, não apresentam significado e nem descrevem um contexto ou cenário. A primeira camada da Figura 1.b apresenta um exemplo de observações dos sensores. Os valores são obtidos sem a aplicação de nenhuma semântica aos dados.

Os sensores possuem metadados que descrevem o sensor e o contexto em que está inserido. Por exemplo, alguns metadados descrevem a unidade e tipo de medida, o serviço

que disponibiliza os dados, a geolocalização, entre outros. Estas descrições quando associadas aos dados permitem que os dados evoluam para uma informação. Desta forma, as observações passam a ser caracterizadas com dados que descrevem o seu significado e contexto.

A Figura 1.b apresenta na camada de informação um exemplo de uma observação de um sensor anotada semanticamente por meio de termos de uma ontologia. Essa anotação representa a informação em seu nível de menor abstração, já que descreve apenas o dado observado e o contexto do dispositivo. Entretanto, ainda na camada de informação pode-se estabelecer informações com diferentes níveis de abstrações, como por exemplo, o valor médio das observações de um sensor de temperatura. Sendo assim, esse valor médio de temperatura equivale a uma informação que resume um conjunto de dados que foi definido a partir de um intervalo de tempo especificado.

A próxima camada diz respeito ao conhecimento. Após as informações serem analisadas ou processadas pode-se deduzir algo em relação a estas. Essa dedução ou conclusão obtida representa a camada de conhecimento. Por exemplo, na Figura 1.b um evento foi obtido como “Tempo Frio”. Este evento pode ser adquirido a partir de uma análise efetuada sobre as informações disponibilizadas, como observações de sensores de temperatura, umidade, intensidade do vento, etc. Com base nas informações analisadas pode-se obter a conclusão de que o tempo está frio ou estará frio em determinado período.

Com essa informação disponibilizada pode-se obter a sabedoria. A camada de sabedoria refere-se às ações e decisões do usuário ou de um sistema mediante o conhecimento adquirido. Por exemplo, se o usuário toma ciência que o tempo está frio em uma determinada cidade para a qual irá viajar, pode decidir incluir roupas mais quentes em sua bagagem. Outro exemplo, suponha que um sistema de uma casa inteligente possua informações relacionadas ao tempo do ambiente externo à casa, sabendo que o tempo está frio, pode acionar um evento para ligar o aquecedor na casa.

3. DIKW4IoT: Construindo um grafo de conhecimento de dados *IoT* baseado na hierarquia *DIKW*

O grafo de conhecimento consiste em uma grande rede de entidades interconectadas. A sua construção se dá a partir da integração de diferentes tipos de conhecimento [Guo et al. 2020]. No contexto de *IoT*, o grafo de conhecimento é estruturado a partir da integração de diferentes sensores e/ou atuadores de um ambiente.

Para [Xie et al. 2020] um grafo de conhecimento de dados *IoT* é composto por uma ontologia de base *IoT*, uma ontologia de domínio *IoT* e uma instância de *IoT*. A ontologia base de *IoT* consiste no vocabulário que será utilizado para padronizar os dispositivos *IoT* e suas observações. A ontologia de domínio *IoT*, por sua vez, tem como proposta estruturar os dispositivos, relações e atributos em um domínio específico. A instância de *IoT* consiste nos dispositivos e dados que serão integrados.

Propõe-se então uma abordagem denominada como DIKW4IoT, em que o grafo de conhecimento com dados de sensores é construído levando em consideração as camadas de dados da hierarquia *DIKW*. A Figura 2 apresenta as etapas necessárias para sua construção.

A etapa **Anotação Semântica de Sensores** consiste na identificação dos sensores

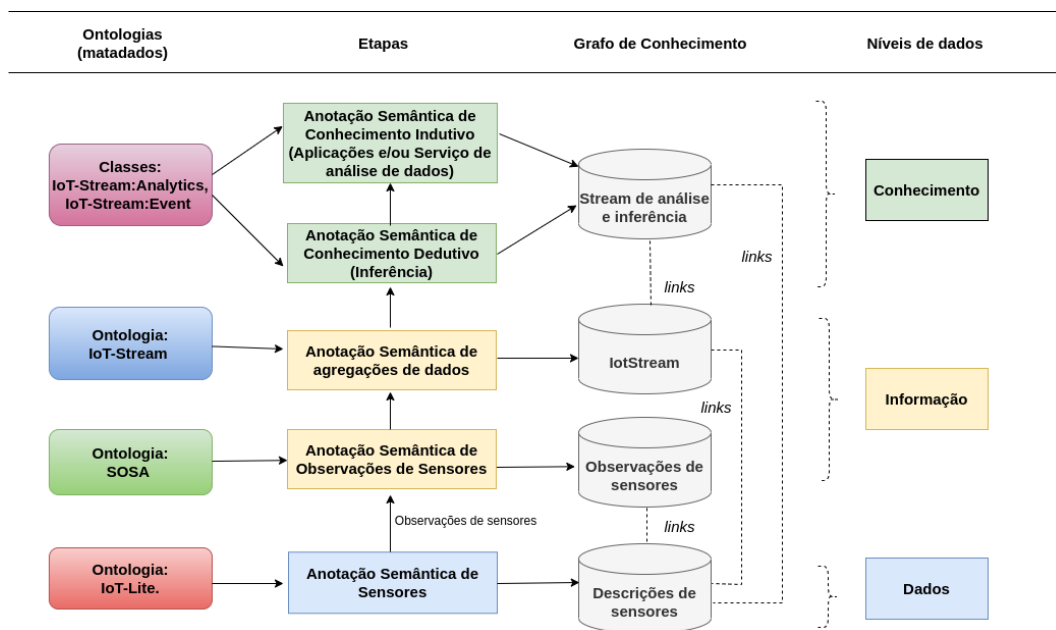


Figura 2. Etapas para construção do grafo de conhecimento de dados IoT baseado na hierarquia DIKW - Fonte: Elaborado pelo autor

que serão integrados no grafo de conhecimento. Nesta etapa são anotados semanticamente dados que descrevem os sensores. Essa etapa é realizada uma única vez para cada sensor. A anotação semântica é realizada por meio da ontologia *IoT-Lite*¹. Essa ontologia possui conceitos e relações que permitem a descrição de sensores. A ontologia *IoT-Lite* reusa termos da ontologia *QUDV* que permitem identificar as unidades de medições utilizadas.

Os sensores cadastrados produzem dados que são chamados de observações. Essas observações correspondem a entrada para a etapa **Anotação Semântica de Observações de Sensores**. Nesta, as observações são anotadas semanticamente por meio de conceitos da ontologia *SOSA*².

A ontologia *SOSA* adota uma perspectiva centrada em eventos e fornece uma especificação de propósito geral abordando conceitos relacionados a observação, atuação, amostragem e procedimentos. Uma observação possui um resultado representado pela propriedade *hasSimpleResult* e uma anotação temporal representada pela propriedade *hasResultTime*. As observações mantêm um vínculo com sua fonte *IoT* por meio da propriedade *madeBySensor*. As observações semânticas no grafo de conhecimento fazem parte da camada de Informação.

A etapa **Anotação Semântica de Agregações de Dados** corresponde a anotação de um resultado proveniente de uma função de agregação. Funções como média, contagem, soma, máximo e mínimo definem uma informação sobre um conjunto de observações. Essas funções são efetuadas de acordo com intervalo de tempo especificado. Logo, o intervalo de tempo define quais observações serão incluídas para o cálculo de agregação.

Por exemplo, a média da intensidade do vento nas últimas 06 horas. O in-

¹<https://www.w3.org/Submission/2015/SUBM-iot-lite-20151126/>

²<https://www.w3.org/TR/vocab-ssn/>

intervalo 06 horas definido por um *timestamp* inicial e por um *timestamp* final definem quais observações do sensor que medem a intensidade do vento farão parte da função de agregação média de valores. Este novo resultado consiste em uma informação mais abstrata que foi obtida sobre um conjunto de observações.

Desta forma, tem-se que os dados que representam a camada de informação podem apresentar diferentes abstrações. Essas abstrações variam de acordo com a especificidade e raciocínio atribuído. Por exemplo, a seguinte afirmação “a temperatura em São Paulo foi 15° C no dia 27 de julho de 2022 às 13h”, tem uma abstração menor e especificidade maior do que afirmação “a maior temperatura do mês de julho em São Paulo foi 17 °C”. Essa última afirmação é mais abstrata e menos específica pois resume um conjunto de dados. Entretanto, não deixa de ser uma informação.

A ontologia *IoT-Stream*³ é utilizada na literatura para anotação de fluxos de dados de forma mais leve. Optou-se por utilizar conceitos da ontologia *IoT-Stream* nesta etapa por possuir propriedades que permitem a descrição do método e do intervalo de agregação que está sendo aplicado. As propriedades são *methods*, que permite identificar o método utilizado sobre as observações, *windowStart* e *windowEnd* que permitem identificar o intervalo aplicado sobre as observações.

A etapa **Anotação Semântica de Conhecimento Dedutivo** correspondem as triplas que são geradas a partir da inferência de axiomas previamente estabelecidos. Os axiomas consistem em regras de lógica de primeira ordem que são especificadas a partir de termos de uma ontologia de domínio e que definem uma verdade sobre o conjunto de dados.

Por exemplo, suponha que em um prédio existe dois tipos de sensores: um sensor para detectar temperatura e um sensor para detectar fumaça. Foi estabelecido que na presença de fumaça no ambiente e temperatura igual ou maior que 37° C indicaria a ocorrência de um “Incêndio”. Na ontologia de domínio foi definida a seguinte regra de axioma: $Fire \equiv (Temperature \sqcup (hasTemperature \geq 37)) \sqcup (Smoke \sqcup (hasSmokePresence true))$. Quando efetuado o processo de raciocínio sobre os dados, um novo fato poderá ser identificado, caso corresponda aos padrões de equivalência estabelecidos. Esse novo fato ou conclusão deduzida é definido como conhecimento.

A etapa **Anotação Semântica de Conhecimento Indutivo** refere-se a anotação semântica dos resultados que podem ser obtidos em serviços de análise de dados e/ou aplicações para detecção de eventos. A anotação destes novos dados estendem as classes *Analytics* e *Event* da ontologia *IoT-Stream* para incluírem conceitos relacionados a ontologia da aplicação.

Por exemplo, em um cenário de Saúde Inteligente, suponha que um paciente é monitorado por sensores capazes de detectar os valores referentes a pressão arterial sistêmica, glicose, oximetria, frequência cardíaca, etc. Uma aplicação pode ser modelada para detectar quando um paciente está prestes a ter um ataque cardíaco de acordo com parâmetros especificados. O evento detectado deve ser anotado semanticamente e adicionado ao grafo de conhecimento. Esse evento é classificado como um dado de Conhecimento, pois é resultado de um processamento em que parâmetros foram avaliados e validados para que o evento fosse detectado.

³<http://iot.ee.surrey.ac.uk/iot-crawler/ontology/iot-stream/>

Quando o conhecimento obtido por meio de aplicações e/ou serviços é compartilhado e anotado semanticamente por meio de um vocabulário comum, o conhecimento deixa de ser restrito a uma aplicação/serviço e passa a ser acessível a todos que tem interesse no conhecimento gerado.

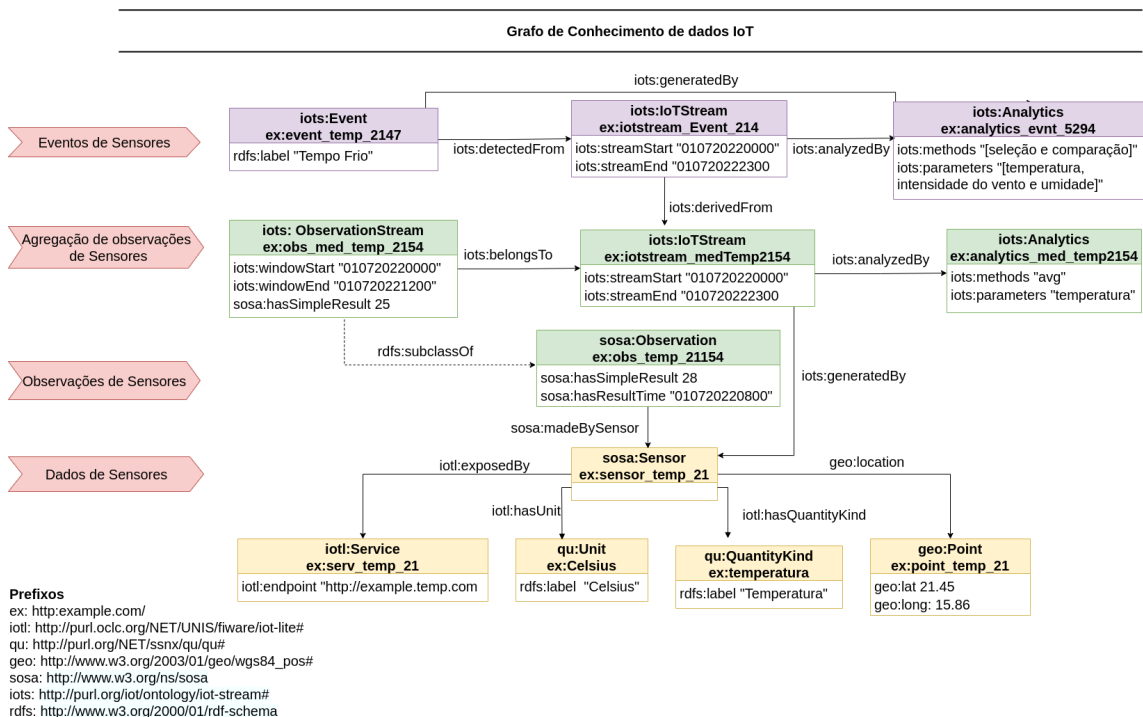


Figura 3. Anotação Semântica dos dados nas diferentes camadas dos dados
 - Fonte: Elaborado pelo autor

A Figura 3 apresenta um exemplo da anotação semântica aplicada aos dados. Em cada camada de dados foram utilizadas as ontologias indicadas na abordagem DIKW4IoT. Perceba que as ontologias utilizadas relacionam-se entre si por compartilharem alguns conceitos e relações. Isso foi possível porque a ontologia *IoT-Lite* consiste em uma instanciação mais leve da ontologia *SSN*. A ontologia *SOSA*, por sua vez, consiste no módulo central da ontologia *SSN*, todavia atuando de forma independente e mais leve, devido ter menos axiomatização. E a ontologia *IoT-Stream* estende a ontologia *SOSA*.

3.1. Workflow para abordagem DIKW4IoT

Este *workflow* foi elaborado para auxiliar a construção do grafo de conhecimento de dados *IoT* utilizando a abordagem DIKW4IoT. A abordagem é composta por duas fases: *offline* e *online*. A fase *offline* é responsável pela definição dos sensores que serão integrados, os mapeamentos de dados e ontologias que serão aplicadas, etc. A fase *online* é responsável pela aquisição das observações de sensores, dos resultados de análise, etc.

A Figura 4 apresenta um esboço das principais tarefas executadas na fase *offline*. A fase *offline* é composta pelos seguintes segmentos: Contexto, Mapeamento de dados, Agregação, Ferramentas e Anotação semântica de dados de sensores.

A definição do contexto inclui identificar os sensores que serão integrados no grafo de conhecimento, assim como definir se o grafo de conhecimento será desenvolvido para

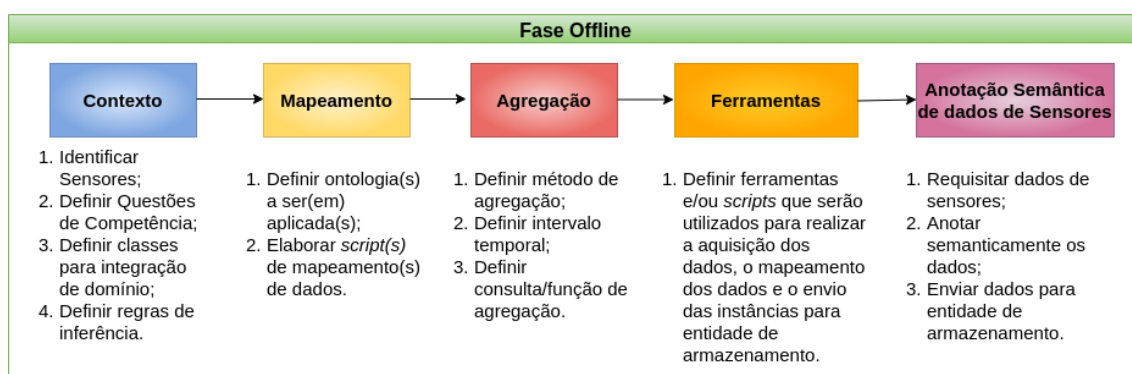


Figura 4. Fase *offline* da abordagem DIKW4IoT - Fonte: Elaborado pelo autor

um domínio específico. Caso seja definido para um domínio específico, deve-se definir como os dados dos sensores serão vinculados à ontologia de domínio.

Na fase *offline* são definidas as ontologias e mapeamentos de dados que serão utilizados na anotação semântica dos dados, assim como as ferramentas necessárias para realizar a aquisição dos dados, o processo de anotação semântica e o envio das *triplas* para o *triplestore*. Nesta fase também é definido o método de agregação que será aplicado sobre os dados para que informações abstratas possam ser construídas.

A anotação semântica dos sensores constitui em requisitar e anotar os dados que descrevem os sensores que serão integrados. As classes e relações que compõem essa anotação definem o esquema inicial do grafo de conhecimento *IoT*.

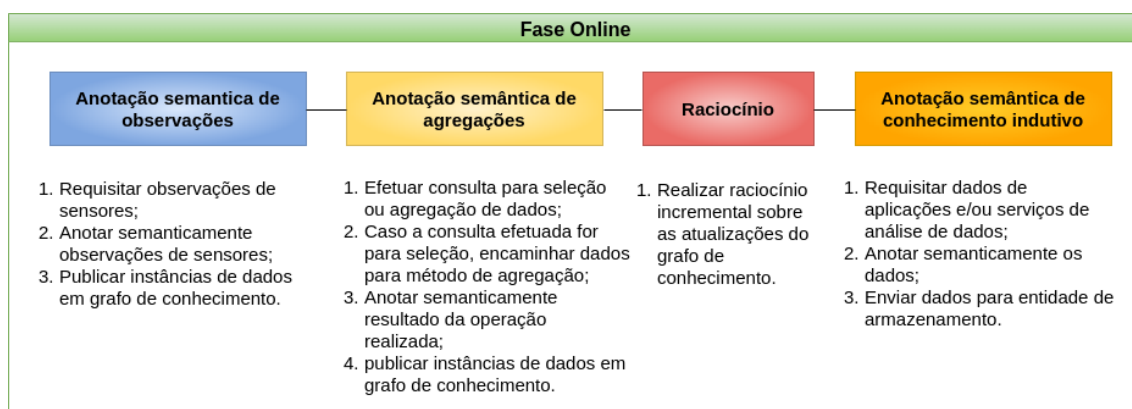


Figura 5. Fase *online* da abordagem DIKW4IoT - Fonte: Elaborado pelo autor

A Figura 5 apresenta as principais tarefas efetuadas na fase *online*. A **fase *online*** é definida pela obtenção dos dados próximo ao tempo real, ou seja, a medida que o serviço de dados disponibiliza os dados referente as observações de sensores, estes são obtidos, anotados semanticamente e publicados no grafo de conhecimento. A proposta desta abordagem é que o grafo de conhecimento de dados *IoT* possua diferentes níveis de dados, sendo assim, outros processamentos serão efetuados sobre os dados.

Na fase *online* é efetuada a aquisição e anotação das observações dos sensores. É realizada também a anotação semântica das agregações que foram definidas anteriormente. Para isso a consulta e execução do método deverão ser realizadas para que o

resultado gerado seja anotado semanticamente.

As tarefas de raciocínio e anotação de conhecimento indutivo referem-se à anotação semântica dos dados para a camada de conhecimento. O conhecimento dedutivo é obtido a partir do processo de raciocínio efetuado sobre os dados. O raciocínio incremental é o mais indicado para realização, isto é, o raciocínio deverá ser efetuado tomando como base as atualizações realizadas no grafo de conhecimento. Este procedimento é indicado devido o processo de raciocínio sobre os dados ser complexo e custoso.

A anotação semântica do conhecimento indutivo deve ser realizado a medida que os serviços e aplicações compartilham seus dados. Estes deverão ser adquiridos, anotados e adicionados ao grafo de conhecimento.

4. Estudo de caso: Construindo um grafo de conhecimento com parâmetros ambientais da cidade de Lisboa

Utilizando-se a abordagem DIKW4IoT, construiu-se um grafo de conhecimento de dados *IoT* com os dados disponibilizados no portal Lisboa aberta⁴. O conjunto de dados é composto por observações de 717 sensores que estão distribuídos em mais de 80 ruas de Lisboa. As observações constituem dados de temperatura, umidade, intensidade do vento, dados referente a qualidade do ar como PM10, PM25, SO₂, NO₂ e etc.

Iniciando a fase *offline* foram identificados e definidos os sensores para integração no grafo de conhecimento. As questões de competência também foram definidas. As questões de competência são importantes para a validação do grafo de conhecimento. Devem evidenciar o que deseja ser respondido através do grafo de conhecimento.

Para a construção do grafo de conhecimento foi desenvolvido um programa⁵ na linguagem Java para realizar a requisição *HTTP* dos dados, anotação semântica e envio de dados para um *triplestore*. O *triplestore* escolhido foi o *GraphDB*⁶. A sua escolha foi devido este possuir uma versão gratuita que disponibiliza diferentes funcionalidades para seleção e visualização dos dados. Os mapeamentos dos dados foram definidos por meio de métodos. O programa adota o padrão de projeto *Adapter*, logo para cada mapeamento de dado foi criado um adaptador.

Após os mapeamentos serem definidos, os dados de sensores foram adquiridos e anotados semanticamente. Na fase *online* foram obtidas observações de sensores no período de 01 de abril de 2022 a 14 de maio de 2022.

Foi adicionado ao grafo de conhecimento uma camada de dados agregados. A agregação foi realizada aplicando o método média sobre as observações de sensores de temperatura, intensidade do vento, umidade, PM10, PM25, O₃, entre outros. Por exemplo, dentre os 717 sensores, 80 são sensores de temperatura distribuídos pela cidade de Lisboa. A agregação reúne as observações dos 80 sensores de temperatura a cada hora e resulta no valor da temperatura média da cidade de Lisboa como um todo. Desta forma, o grafo dispõe o valor da temperatura de Lisboa em pontos específicos, assim como dispõe da média geral de temperatura para a cidade.

⁴<https://lisboaaberta.cm-lisboa.pt/index.php/pt/dados/conjuntos-de-dados>

⁵<https://github.com/searchmdcc/frameworkGraphThings>

⁶<https://graphdb.ontotext.com/>

A fim de demonstrar o consumo dos dados do grafo de conhecimento criado, elaborou-se um aplicativo *mobile* para apresentação de informações relacionadas ao tempo e a qualidade do ar em Lisboa, assim como a detecção de eventos mediante o consumo do grafo de conhecimento. Os eventos detectados são enviados para publicação no grafo como representação dos dados da camada de conhecimento.

O aplicativo *AppMonitoraLisboa*⁷ foi desenvolvido utilizando a plataforma *android*⁸. Tem como funções básicas fornecer informações gerais sobre o tempo, por exemplo, temperatura, umidade e intensidade do vento, informações relacionadas a qualidade do ar e sua classificação em bom, médio e ruim.

Para acesso ao grafo de conhecimento utilizou-se algumas bibliotecas do *framework Jena* para *android*⁹. O *framework Jena* oferece mecanismos para consultar um grafo de conhecimento por meio de consultas *Sparql*¹⁰. Todavia, em sua versão para *android* possui algumas limitações quando comparada com a versão *Jena*¹¹ para *java*.

A Figura 6 apresenta um exemplo da aplicação. A Figura 6.A apresenta informações da média da temperatura, umidade e intensidade do vento dos respectivos sensores distribuídos na cidade de Lisboa.

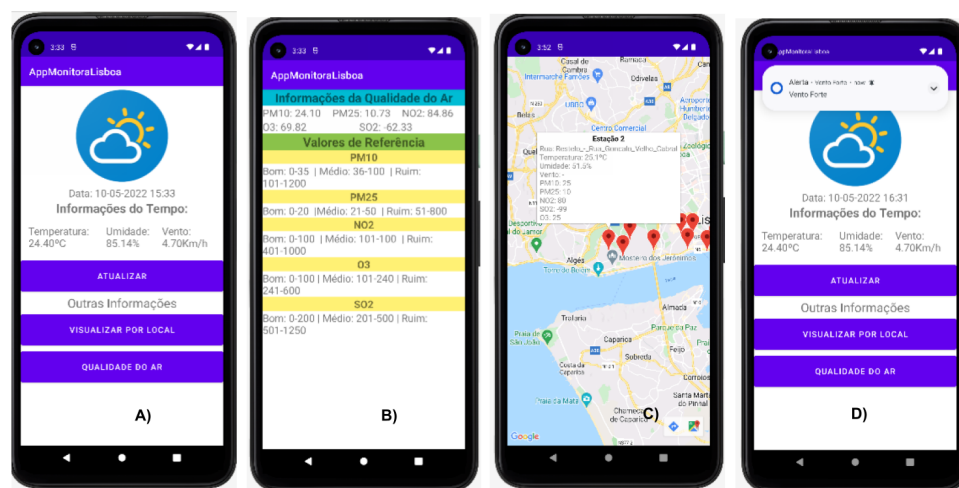


Figura 6. AppMonitoraLisboa - Fonte: elaborado pelo autor

A Figura 6.B apresenta a média dos valores dos sensores de PM10, PM25, SO2, NO2 e O3. Estes índices são utilizados para avaliar a qualidade do ar. Como pode ser visualizado na Figura 6.B é disponibilizado alguns valores de referência que auxiliam o usuário a classificar a qualidade do ar no momento.

A Figura 6.C permite ao usuário visualizar as estações de sensores. Ao clicar em cada marcação é apresentada os resultados dos sensores presentes naquela estação. Desta forma, o usuário pode visualizar informações mais específicas de uma determinada área de Lisboa.

⁷<https://github.com/searchmdcc/appMonitoraLisboa>

⁸<https://developer.android.com/docs?hl=pt-br>

⁹<https://github.com/maddymz/Topic-Recommendation-System/tree/master/app/app/libs>

¹⁰<https://www.w3.org/TR/sparql11-query/>

¹¹<https://jena.apache.org/>

A aplicação possui a funcionalidade de monitoramento de eventos. Alguns eventos simples foram configurados como “Vento Forte” para intensidade do vento maior que 70 km/h, “Baixa Umidade” para valores de umidade abaixo de 30%, entre outros eventos. Ao detectar um dos eventos estabelecidos uma notificação é enviada para o usuário, como pode ser visualizada na Figura 6.D. Este evento é enviado para uma base de dados no *Firebase*¹². Posteriormente, este evento é adquirido, anotado semanticamente e compartilhado no grafo de conhecimento.

5. Trabalhos Relacionados

A estrutura DIKW4IoT tem como proposta criar uma base de conhecimento que contemple os diferentes níveis de dados abordados pela hierarquia *DIKW* [Van Meter 2020]. Baseada em uma hierarquia centrada em dados, apresenta quatro camadas de dados que utilizam diferentes ontologias em sua representação.

A primeira e segunda camada de dados referem-se a anotação semântica dos dados de sensores e suas observações. Para isso utiliza as ontologias *IoT-Lite* e *Sosa*, respectivamente. A terceira e quarta camada referem-se a anotação semântica de resultados de agregações de dados e resultados provindos de serviços de análises de dados ou eventos. Para isso é utilizado a ontologia *IoT-Stream*.

Trabalhos	Dados de sensores	Observações de sensores	Resultados de agregação de dados	Resultados de análise de dados e eventos
[Le-Phuoc et al. 2016]	✓	✓	-	-
[Ding et al. 2020]	✓	✓	-	-
[Soares et al. 2019]	✓	✓	-	-
[Alvarez-Coello and Gómez 2021]	-	-	✓	✓
DIKW4IoT	✓	✓	✓	✓

Tabela 1. Análise de trabalhos relacionados baseado nas camadas de DIKW4IoT para construção de base de conhecimento IoT

A pesquisa abordada em [Le-Phuoc et al. 2016] apresenta a construção de um grafo de conhecimento com dados de sensores. Este grafo possui apenas dados referentes as camadas relacionadas aos dados de sensores e observações. Para anotação semântica é utilizada a ontologia *SSN (Semantic Sensor Network Ontology)*¹³.

No trabalho de [Soares et al. 2019] é abordado a integração de dados relacionados a um domínio marítimo. Para isso utiliza as fontes de dados *AIS* e *IMO vessels*, com informações relacionadas a embarcações e dados sobre estações climáticas e estações de gelo. Utiliza a ontologia *SSN* e uma ontologia de domínio para anotação semântica dos dados. O grafo construído não contempla resultados de agregações e análises de dados.

O trabalho de [Ding et al. 2020] apresenta um *framework* baseado em ontologia para integração e análise de dados. Utiliza a ontologia *SSN* e a ontologia *geoSparql* para

¹²<https://firebase.google.com/>

¹³<https://www.w3.org/TR/vocab-ssn/>

anotação semântica dos dados. A integração dos dados adota a abordagem virtual, isto é, não realiza a materialização dos dados e tem como foco apenas dados de sensores e observações.

A pesquisa de [Alvarez-Coello and Gómez 2021], por sua vez, realiza a integração de dados no domínio automotivo. Diferente dos demais trabalhos, não realiza a anotação semântica de sensores e observações. Tem como proposta a anotação semântica dos resultados de uma análise de dados, ou seja, realiza-se a anotação do resultado derivado de uma média, desvio padrão e soma dos valores das observações de um sensor dentro de uma janela de tempo instituída. Realiza também a anotação de eventos relacionados a mudanças que ocorrem em um fluxo categórico. Utiliza uma abordagem denominada como *top-down*, ou seja, a integração dos dados é realizada inicialmente a partir dos dados resultantes de serviços de análises e aplicações. Essa integração equivale a terceira e quarta camada proposta por DIKW4IoT.

Para um melhor entendimento elaborou-se a tabela 1 para comparação dos trabalhos tendo como base as etapas abordadas por DIKW4IoT para construção de uma base de conhecimento *IoT*. Perceba que os trabalhos abordados não contemplam as diferentes camadas de dados da hierarquia *DIKW* em sua totalidade. Ao contrário da abordagem DIKW4IoT que tem como objetivo a construção da base de conhecimento *IoT* centrada nesta hierarquia de dados.

6. Conclusão e trabalhos futuros

Este trabalho discorreu sobre uma abordagem para a construção de um grafo de conhecimento de dados *IoT* prezando pela representação dos dados baseado na hierarquia *DIKW*. Para isso foram estabelecidas as relações sobre os diferentes níveis de dados e como eles podem ser representados no grafo de conhecimento. A abordagem é denominada DIKW4IoT e utiliza conceitos e relações das ontologias *IoT-Lite*, *SOSA* e *IoT-Stream* para anotação semântica dos diferentes níveis de dados.

Elucidou-se a importância de realizar a realimentação do grafo de conhecimento com os dados resultantes de serviços de análises e/ou aplicações. Estes dados representam no grafo de conhecimento o equivalente a camada de conhecimento da hierarquia *DIKW*. Compartilhar este tipo de informação mais abstrata e/ou especializada contribui para que serviços mais especializados possam ser desenvolvidos, assim como amplia a visão geral para a tomada de decisão.

Como prova de validação da abordagem utilizada elaborou-se um grafo de conhecimento com dados referentes a parâmetros ambientais da cidade de Lisboa e uma aplicação *mobile* para consumir e gerar dados de eventos para serem publicados no grafo de conhecimento.

Em trabalhos futuros objetiva-se validar o uso do *workflow* proposto analisando a sequência dos passos efetuados, o desempenho em cada tarefa e o tempo para execução, comparando com outras estruturas para construção de grafo de conhecimento que não utilizam os passos sequenciados.

Referências

Alvarez-Coello, D. and Gómez, J. M. (2021). Ontology-based integration of vehicle-related data. In *2021 IEEE 15th International Conference on Semantic Computing*

- (ICSC), pages 437–442. IEEE.
- Ding, L., Xiao, G., Calvanese, D., and Meng, L. (2020). A framework uniting ontology-based geodata integration and geovisual analytics. *ISPRS International Journal of Geo-Information*, 9(8):474.
- Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q., and Lin, Z. (2017). Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. In *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 327–332. IEEE.
- Gao, H., Duan, Y., Shao, L., and Sun, X. (2021). Transformation-based processing of typed resources for multimedia sources in the iot environment. *Wireless Networks*, 27(5):3377–3393.
- Guo, S., Niu, G., Wang, Z., Pun, M.-O., and Yang, K. (2020). An indoor knowledge graph framework for efficient pedestrian localization. *IEEE Sensors Journal*, 21(4):5151–5163.
- Gyrard, A., Patel, P., Sheth, A. P., and Serrano, M. (2016). Building the web of knowledge with smart iot applications. *IEEE Intelligent Systems*, 31(5):83.
- Le-Phuoc, D., Quoc, H. N. M., Quoc, H. N., Nhat, T. T., and Hauswirth, M. (2016). The graph of things: A step towards the live knowledge graph of connected things. *Journal of Web Semantics*, 37:25–35.
- Liew, A. (2013). Dikiw: Data, information, knowledge, intelligence, wisdom and their interrelationships. *Business Management Dynamics*, 2(10):49.
- Pan, J. Z., Vetere, G., Gomez-Perez, J. M., and Wu, H. (2017). *Exploiting linked data and knowledge graphs in large organisations*. Springer.
- Remor, C. A., Fialho, F. A., and Queiroz, M. P. (2017). Analisando a hierarquia dikw. In *Anais do Congresso Internacional de Conhecimento e Inovação–ciki*, volume 1.
- Roberto, L. (2021). A pirâmide do conhecimento. Artigo disponível em <https://professorluizroberto.com/a-piramide-do-conhecimento/>. Acesso: 07/04/2022.
- Soares, A., Dividino, R., Abreu, F., Brousseau, M., Isenor, A. W., Webb, S., and Matwin, S. (2019). Crisis: Integrating ais and ocean data streams using semantic web standards for event detection. In *2019 International conference on military communications and information systems (ICMCIS)*, pages 1–7. IEEE.
- Steindl, G., Heinzl, B., Kastner, W., and Mauri, J. (2019). A novel ontology-based smart service architecture for data-driven model development. In *eKNOW 2019-The Eleventh International Conference on Information, Process, and Knowledge Management*, pages 26–27.
- Van Meter, H. J. (2020). Revising the dikw pyramid and the real relationship between data, information, knowledge, and wisdom. *Law, Technology and Humans*, 2(2):69–80.
- Xie, C., Yu, B., Zeng, Z., Yang, Y., and Liu, Q. (2020). Multilayer internet-of-things middleware based on knowledge graph. *IEEE Internet of Things Journal*, 8(4):2635–2648.