Chapter 10
Case Study from the Energy Domain

Dea Pujić, Marko Jelić, Nikola Tomašević, and Marko Batić

Institute Mihajlo Pupin
University of Belgrade, Serbia

Abstract. Information systems are most often focused on when considering applications of Big Data technology. However, the energy domain is more than suitable also given the worldwide coverage of electrification. Additionally, the energy sector has been recognized to be in dire need of modernization, which would include tackling (i.e. processing, storing and interpreting) a vast amount of data. The motivation for including a case study on the applications of big data technologies in the energy domain is clear, and is thus the purpose of this chapter. An application of linked data and post-processing energy data has been covered, whilst a special focus has been put on the analytical services involved, concrete methodologies and their exploitation.

1 Introduction

Big Data technologies are often used in domains where data is generated, stored and processed at rates that cannot be efficiently processed by one computer. One of those domains is definitely that of energy. Here, the processes of energy generation, transmission, distribution and use have to be concurrently monitored and analyzed in order to assure system stability without brownouts or blackouts. The transmission systems (grids) that transport electric energy are in general very large and robust infrastructures that are accompanied by a great deal of monitoring equipment. Novel Internet of Things (IoT) concepts of smart and interconnected homes are also pushing both sensors and actuators into peoples homes. The power supply of any country is considered to be one the most critical systems and as such its stability is of utmost importance. To that effect, a wide variety of systems are deployed for monitoring and control. Some of these tools are presented in this chapter with a few from the perspective of end users (Non-Intrusive Load Monitoring, Energy Conservation Measures and User Benchmarking) and a few from the perspective of the grid (production, demand and price forecasting).

2 Challenges withing the Big Data Energy Domain

In order to be able to provide advanced smart grid, user-oriented services, which will be discussed further in this chapter, integration with high volume, heterogeneous smart metering data (coming both from the grid side, e.g. placed in power substations, and from the user side, e.g. installed in homes and buildings) is a prerequisite. To specify, suggest and deliver adequate services to end users (i.e. energy consumers) with respect to their requirements and power grid status, various forms of energy data analytics should be applied by distribution system operators (DSO) and grid operators such as precise short- and long-term energy production and consumption forecasting. In order to deliver such energy analytics, historical energy production data from renewable energy sources (RES) and historical consumption data, based on smart metering at consumer premises and LV/MV power substations, must be taken into account.

The main challenge to providing advanced smart grid services is related to the integration and interoperability of high volume heterogeneous data sources as well as adequate processing of the acquired data. Furthermore, making this data interoperable, based on Linked Data API, and interlinked with other data sources, such as weather data for renewable energy sources (RET) production analysis, number of inhabitants per home units, etc., is essential for providing additional efficient user tailored analytical services such as energy conservation action suggestions, comparison with other consumers of the same type, etc.

Another challenge is related to analysis of grid operations, fault diagnostics and detection. To provide such advanced analytics, real-time integration and big data analysis performed upon the high volume data streams coming from metering devices and power grid elements (e.g. switches, transformers, etc.) is necessary, and could be solved using Linked Data principles. Finally, to support next generation technologies enabling smart grids with an increased share of renewables, it is necessary to provide highly modular and adaptable power grids. In addition, adequate tools for off-line analysis of power system optimal design should be deployed. These analytical tools should also incorporate allocation of optimal reconfiguration of power grid elements to provide reliable and flexible operation as an answer to the changing operational conditions. Tools for planning and reconfiguring power distribution networks consider power station infrastructure and its design, number and capacity of power lines, etc. To provide such advanced grid capabilities, integration with historical power grid data, archives of detected alarms and other relevant operational data (such as data from smart metering, consumption data, etc.) is necessary. Therefore, the main challenge is to provide digested input to the batch-processing, big data analytics for power grid infrastructure planning.

Having all of this in mind, the significance of big data processing techniques is obvious. On the other hand, further in this chapter examples of analytical services will be presented and discussed.

3 Energy Conservation Big Data Analytical Services

Improving quality of life through advanced analytics is common nowadays in various domains. Consequently, within the energy domain, collecting data from numerous smart meters, processing it and drawing conclusions are common concepts in the field of developing energy conversation services. The amount of aforementioned data highly depends on the service's principal use. If the focus is put on just one household, data can be undoubtedly processed using only one computer. Nonetheless, if the scale of a problem is a neighbourhood, municipality or city level, data processing and analytical computations can be taken as a big data problem. Therefore, within this chapter, methodologies for smart energy services are going to be discussed.

3.1 Non-Intrusive Load Monitoring

The first of these is so-called Non-Intrusive Load Monitoring (NILM). NILM was motivated by conclusions, such as those from [69], which claimed that up to 12% of residential energy consumption can be decreased by giving users feedback on how the energy has been used. In other words, by providing the user with information about which of their appliances is using electrical energy and how much, significant savings can be reached. Nonetheless, providing this kind of information would require installation of numerous meters all around households, which is usually unacceptable for the end-user. Therefore, instead of the Intrusive Load Monitoring solution which influences users' convenience, Non-Intrusive Load Monitoring was proposed by Hart in [183] with the main goal of providing users with the same information in a harmless way by aggregating entire household consumption at the appliance level, which can be seen in 1.

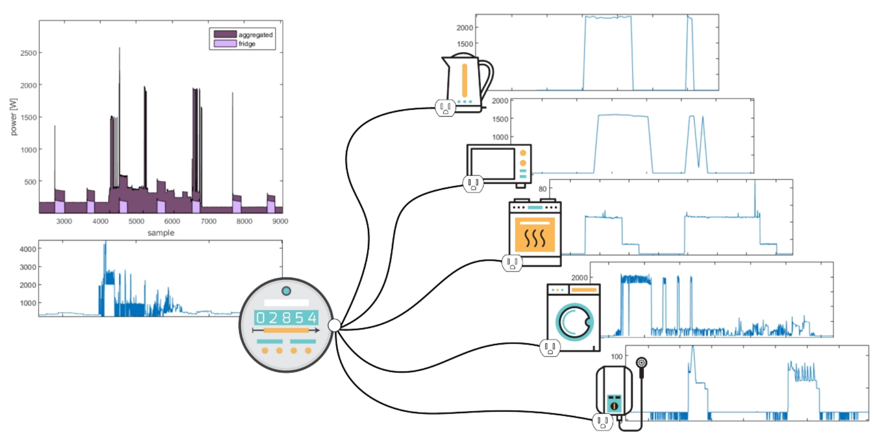

Fig. 1: Non-Intrusive Load Monitoring concept

Having in mind the previous information, two main problems are present within the NILM literature - **classification**, which provides information about the activation on the appliance level, and **regression** for the estimation of the appliance's individual consumption, as shown in the example Figure 2. As these are some of the most common problems in advanced analytics, typical methodologies employed to address these are leading machine learning approaches, which are going to be presented and discussed further in this section to give an example of the use of applied big data technologies in the energy domain.

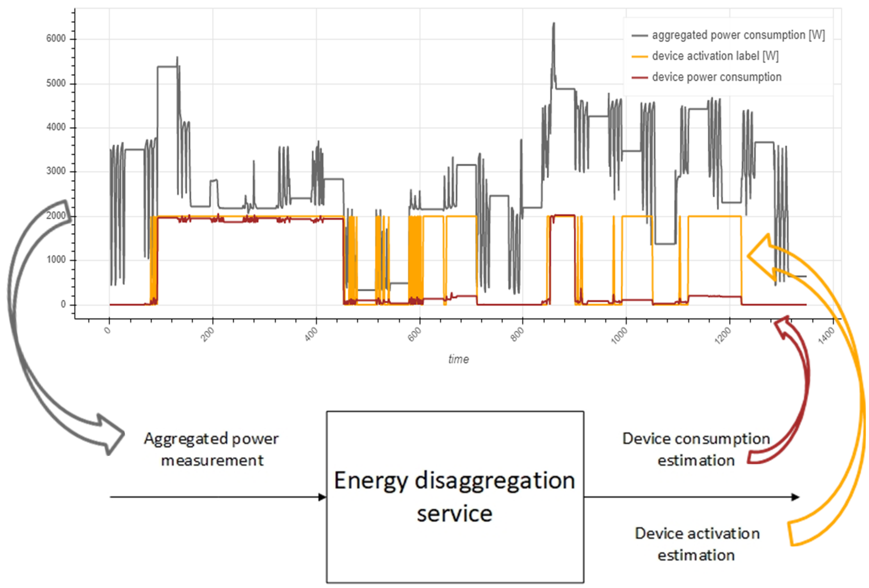

Fig. 2: NILM classification and regression example

As a first step, in this section, the currently present publicly available datasets will be introduced as the basis of data-driven models, which will be discussed further. Depending on the sampling rate, within the NILM literature, data and further corresponding methodologies are usually separated in two groups - **high** and **low** frequency ones. For high frequency, measurements with a sampling time of less than 1 ms are considered. These kind of data are usually unavailable in everyday practice due to the fact that usual residential metering equipment has a sampling period around 1 s and is put as the low frequency group. This difference in sampling rate further influences the choice of the disaggregation methodology and preprocessing approach for the real-time coming data used as the corresponding inputs.

When discussing publicly available data sets, methodologies are not strictly separated in accordance with the chosen sampling rate but rather by the geographical location. In other words, measurements usually correspond to some localized neighbourhood from which both high and low frequency data might be found in the same data set. The first published dataset we refer to is REDD (Reference Energy Disaggregation Data Set, 2011) [255]. It includes both low and high sampling frequency measurements from six homes in the USA. For the first group, both individual and aggregated power measurements were covered for 16 different appliances, allowing the development of various models, which require labeled data. By contrast, high frequency measurements contain only aggregated data from the household, so the developers have to use unsupervised techniques. Another widely spread and used data set published with [237] is UK-DALE (UK Domestic Appliance-Level Electricity) collected in the United Kingdom from five houses. It, again, covers the whole range of sampling rates, and, similarly to REDD, contains labeled data only for those with a sampling period bigger than 1 s. Additional data sets that should be addressed are REFIT [316], ECO (Electricity Consumption and Occupancy) [32], IHEPCDS (Individual household electric power consumption Data Set) [317] for low sampling rate and BLUED [136] and PLAID [144] for the high one[1].

After presenting the available data, potential and common problems with data processing as part of the theme of big data will be discussed. The first one, present in most of the data sets, is the presence of the **missing data**. Depending on the data set and the specific household appliance, the scale of this problem varies. For example, in the case of refrigerators, this is a minor problem which can be neglected because it works circularly, so each approximately 20 min it turns on or off, leading to numerous examples of both active and inactive working periods. By contrast, when, for example, a washing machine is considered, dropping down the sequence of its activation is unacceptable as it is turned on twice a week in a household on average, so it is difficult to collect enough data for training purposes. Therefore, different techniques were adapted in different papers for additional data synthesization from simply adding existing individual measurements of the appliance's consumption on the aggregated power measurements in some intervals when the considered appliance has not been working to more sophisticated approaches such as generative modeling, which was used to enrich data from commercial sector measurements [192].

It is worth mentioning here that characteristics of the data from these different sets significantly deviate in some aspects as a result of differences in location, habits, choice of domestic appliance, number of occupants, the average age of the occupant etc. The NILM literature has attempted to address this **generalization problem**. Even though the problem of achieving as high performance as possible on the testing rather than training domain is a hot topic in many fields of research within Machine Learning (ML) and Big Data, the generalization problem is even more crucial for NILM. As different houses might include different types of the same appliances, the performance on the data coming from the house whose measurements have not been used in the training process might be significantly lower than the estimated one. Additionally, it is obvious that the only application of the NILM models would be in houses which have not been used in the training phase, as they do not have labeled data (otherwise, there would be no need for NILM). Bearing all of this in mind, validating the results from the data coming from the house whose measurements have already been used in the training process is considered inadequate. Thus, it is accepted that for validation and testing purposes one, so called, unseen house is set aside and all further validation and testing is done for that specific house. Nonetheless, the

[1] http://wiki.nilm.eu/datasets.html

houses covered by some publicly available dataset are by the rule in the same neighbourhood, which leads to the fact that data-driven models learn patterns which are characteristics of the domain rather than the problem. Therefore, separation of the house from the same dataset might be adequate. Finally, the last option would be validating and testing the measurements from the house using a different data set.

State-of-the-art NILM methodologies will be presented later in this section alongside corresponding estimated performance evaluations. Historically, the first ones were Hidden Markov Models and their advancements. They were designed to model the processes with unobservable states, which is indeed the case with the NILM problem. In other words, the goal is to estimate individual consumption in accordance with the observable output (aggregated consumption). This approach and its improvements have been exploited in numerous papers such as [291], [292], [226], [254], [244], and [55]. However, in all of the previously listed papers which cover the application of numerous HMM advancements to the NILM problem, the problem of error propagation is present. Namely, as HMM presumes that a current state depends on a previous one, mistakes in estimating previous states have a significant influence on predicting current ones.

Apart from HMMs, there are numerous unsupervised techniques applied for NILM. The main cause of this is the fact that labeled data for the houses in which services are going to be installed are not available, as already discussed. Therefore, many authors choose to use unsupervised learning techniques instead of improving generalization on the supervised ones. Examples of these attempts are shown in [193] where clusterization and histogram analysis has been employed before using the conditional random fields approach, in [342] where adaptation over unlabeled data has been carried out in order to improve performance on the gaining houses, and in [135] where disaggregation was described as a single-channel source separation problem and Non Negative Matrix Factorization and Separation Via Tensor and Matrix Factorization were used. Most of these approaches were compared with the HMM-based one and showed significant improvements. Another approach to gain the best generalization capabilities possible that can be found in the literature is semi-supervised concept in which a combination of supervised and unsupervised learning is present. In [29], self-training has been carried out using internal and external information in order to decrease the necessity of labeled data. Further, [207] proposes the application of transfer learning and blind learning, which exploits data from training and testing houses.

Finally, supervised techniques were widely spread in the literature as well. Currently, various ML algorithms hold a prime position with regards to supervised approaches, as they have proven themselves to be an adequate solution for the discussed problem, as reviewed in [417]. The biggest group currently popular in the literature is neural networks (NNs). Their ability to extract complex features from an input sequence was confirmed to increase their final prediction performance. Namely, two groups stood out to be most frequently used - Recurrent Neural Networks (RNNs) with the accent on Long Short Term Memory (LSTM) [300], and Convolutional Neural Networks (CNNs) with a specific subcategory of Denoising Autoencoders [238].

After presenting various analytical approaches for solving the NILM problem, it is crucial to finish this subsection with the conclusion that results obtained by this service could be further post-processed and exploited. Namely, disaggregated consumption at the appliance level could be utilized for developing failure detection services in cooperation with other heterogeneous data.

3.2 Energy Conservation Measures (ECM)

When discussing the appeal and benefits of energy savings and energy conservation amongst end users, especially residential ones, it is no surprise that users react most positively and vocally when potential cost savings are mentioned. Of course, when this is the main focus, retrofitting old technologies, improving insulation materials, replacing windows and installing newer and more energy-efficient technologies is usually included in the course of action first recommended. This is mainly because the aspects that are tackled by these modifications are the largest source of potential heat losses and energy conversion inefficiencies. However, there is a significant and still untapped potential for achieving significant energy savings by correcting some aspects of user behaviour.

Besides inefficient materials, bad habits are one of the main causes of high energy loss, especially in heating and cooling applications with the thermal demand being a distinct issue due to the high volume of energy being spent in the residential sector on it. Finding the crucial behavioral patterns that users exhibit when unnecessarily wasting energy is key for efficient mitigation and, therefore, a smart home concept is proposed in order to analyze user behavior and facilitate the necessary changes. In order to obtain data to be able to suggest energy conservation measures, a set of smart sensors should be deployed to monitor various parameters. Some of these sensors could include but are not limited to:

– Smart external meter interfaces (measurement of total energy consumption in real-time);
– Smart electricity plugs and cables (measurement of energy consumption per appliance in real time and possibility of on/off control);
– Smart thermostats (measurement and continuous control of reference temperature and possibly consumed energy);
– Occupancy sensors (measurement of occupancy and motion and ambient temperature also);
– Window sensors (measurements of open/close status of windows and doors and ambient temperature also);
– Volatile organic compound (VOC) sensors (measurement of air quality and ambient temperature)

In some cases where installing smart plugs and cables is not deemed to be economical, a NILM algorithm described in Subsection 3.1 can be employed in order to infer individual appliance activity statuses using only the data from the external meter. When widespread deployment of such sensors is being done, the amount of data that should be collected, stored and processed quickly grows due to the fact that multiple sensors are to be deployed in each room and that each of the sensors usually reports multiple measurements (e.g. the window sensor reports the temperature besides the open/close status, but also has a set of utility measurements such is the network status strength, battery status, etc. which should also be monitored as they provide crucial data regarding the health of the device itself). Therefore, efficient solutions, possibly from the realm of big data, should be employed in order to facilitate efficient storage and processing of data as the problematic user behavior is time-limited and should be pointed out to the user in due course while a problematic event is ongoing.

A small-scale use case of such a system was tested on around two dozen apartments in the suburbs of Leers, France with the proposed architecture of the system illustrated in Figure 3. Using such an architecture, the back-end of

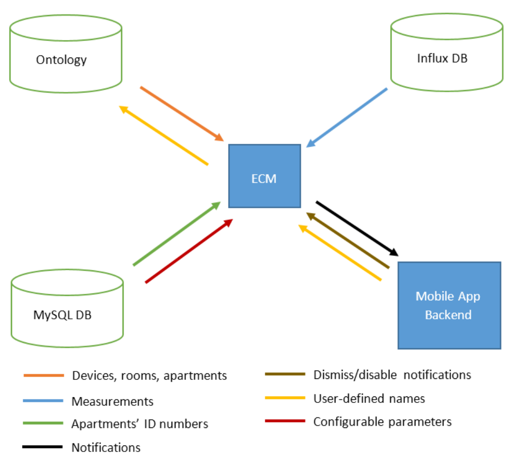

Fig. 3: Proposed architecture of a small-scalle ECM system

the system that employs a MySQL database for static data storage regarding the apartment IDs and custom notification settings in conjunction with an ontology for storing room layouts and detailed sensor deployment data provides support for the main ECM engine that analyses data from the real-time IoT-optimized NoSQL Influx database and sends push notifications to the end users notifying them of energy-inefficient behaviour by cross-correlating different measurements from different sensors. For example, when a heating or cooling device is observed to be turned on in an unoccupied space, the user is warned. If the user acts upon such information and resolves the issue, the notification is dismissed automatically, or if the user does not react and the problematic event goes unresolved, he or she is re-notified after a predefined period of time. These events are analyzed with different scopes for individual rooms but also for entire apartments. Also, since smart sensors are already deployed, the energy conservation analysis can also be extended to regard security (no occupancy whilst a door or window is open) and health (poor air quality and windows closed) aspects also. Of course, each event is analyzed separately and appropriate notifications with corrective actions are issued to the end user.

3.3 User Benchmark

Besides the most obvious motivating factor of energy savings – monetary savings – another factor that can greatly impact users' behavior is social pressure. Namely, in a hypothetical scenario where different users were placed in a competition-like environment where the main goal is to be as energy-efficient as possible or, in other words, where each user's score is determined by how efficiently they consume energy, those users would be more likely to strive to perform better and hence consume energy in a more environmentally friendly way. In order to facilitate such an environment, a benchmarking engine has to be developed in order to provide an algorithm that would rank the users.

[112], [80] and [327] in the literature point out that the benchmarking procedures in the residential sector have long been neglected in favor of industrial applications. Different algorithms and technologies proposed as core include:

- Simple normalization
- Ordinary least squares (OLS)
- Stochastic frontier analysis (SFA)
- Data envelopment analysis (DEA)
- Simulation (model-based) rankings
- Artificial neural networsk (ANNs)
- Fuzzy reasoning

with related literature [171] offering several dozens of additional related algorithms for multi-criteria decision making (MCDM). The applications of the aforementioned algorithms found in the literature are generally focused on schools, other public buildings and offices, with very few papers, such as [289], [258] and [458], analyzing the residential sector.

One of the most prominent standards in energy efficiency ranking is the acclaimed Energy Star program [182], which rates buildings on a scale from 1 to 100 based on models and normalization methods of statistical analysis performed over a database from the US Energy Information Administration (EIA). However, the Energy Star rating does not take into account dynamic data obtained by observing the ongoing behavior of residents. This is where the concept of an IoT-powered smart home can provide a new dimension to energy efficiency benchmarking through real-time analysis of incoming data on how people use the space and appliances at their disposal.

The basis of every ranking algorithm is a set of static parameters that roughly determines the thermal demand of the considered property. These parameters generally include: total heated area, total heated volume, outward wall area, wall thickness, wall conductivity or material, number of reported tenants. This data generally is not massive in volume and is sufficient for some elementary ranking methods. However, an energy efficiency rating that only takes into consideration this data would only have to be calculated once the building is constructed or if some major renovations or retrofits are being made. As such, it would not be able to facilitate a dynamic competition-based environment in which users would compete on a daily or weekly basis on who is consuming their energy in the most economical way.

Given the reasoning above, the static construction and occupancy parameters are extended with a set of dynamic parameters that are inferred based on sensor data collected by the smart home. This data could, for example, include: total consumed energy, occupancy for the entire household, cooling and heating degree days, responsiveness to user-tailored behavior-correcting messages, alignment of load with production from renewable sources, etc. As these parameters are changing on a day-to-day basis, their dynamic nature would provide a fast-paced source that would power the fluctuations in energy efficiency scores of individual users and ultimately help users to see that their change in behaviour has made an impact on their ranking. Also, it is worth mentioning that when users within a same micro-climate are to be ranked, using heating and cooling degree days may prove to be redundant as all users would have the same parameters in this regard. Therefore, this data can be augmented using indoor ambient temperature measurements in order to monitor overheating in winter and overcooling in summer.

The most important procedure that should be conducted within user benchmarking solutions in order to provide a fair comparison between different users with different habits and daily routines is to provide a so-called normalization of consumed energy. This means that, for example, larger consumers should not be discriminated just based on higher consumption; rather, other factors such as the amount of space that requires air conditioning or the number of people using the considered space should be taken into account. In this regard, simply dividing the total consumed energy by the, for example, heated area provides a good first estimate of how energy-efficient different users are per unit of surface, but also implies that a linear relation between area and energy is assumed, which might not be their inherent relationship. In order to mitigate against this issue, vast amounts of data should be collected from individual households using IoT sensors and analyzed in order to either deduce appropriate relations required for normalization or to provide a basis for the aforementioned algorithms (DEA, SFA, etc.), which assign different weights to each of the parameters taken into account.

4 Forecasters

Following the widespread deployment of renewable sources such as wind turbines, photovolotaic panels, geothermal sources, biomass plants, solar thermal collectors and others, mainly as a result of various government-enforced schemes, programs and applicable feed-in tariffs, the stability of the grid has been significantly compromised. The integration of these novel sources has proven to be a relatively cumbersome task due to their stochastic nature and variable production profile, which will be covered in greater depth in Subsection 4.2. Since the production of most of these sources is highly correlated with meteorological data (wind turbine production with wind speed and photovoltaic production with irradiance and cloud coverage), legacy electrical generation capacities (coal, nuclear and hydro power plants) which have a significantly shorter transient between different states of power output have to balance the fast-paced variations in generation that are a byproduct of the introduction of renewable sources. Since total generation is planned in order to be able to fulfill the total demand that will be requested, being able to know beforehand how much energy will be required in the future and how much energy will be available can provide a basis for potential energy and cost savings through optimal resource planning.

4.1 Demand Forecaster

Given the importance of demand forecasting, it is expected that this topic will be covered by more than a few authors in their published research. However, even though there is a noticeable number of publications in this regard, the topic of energy demand forecasting and the methods used for its estimation still appear to be under-explored without a unified proposed approach and most of the studies being case-specific. In that regard, a probabilistic approach for peak demand production is analyzed in [320], an autoregressive model for intra-hour and hourly demand in [447] and ANN-powered short-term forecasting in [399]. Short-term forecasting is also analyzed whilst making use of MARS, SVR and ARIMA models in [9] and [460] presenting a predictive ML approach. Deep learning frameworks are discussed by [33] and [463]. DSM in connection with time-of-use tariffs is analyzed by [199] and simultaneous predictions of electricity price and demand in smart grids in [312].

Some authors like [104], [149], [194] and [12] also discuss demand forecasting but place the focus of their research on the predictors that can be used to predict and correlate with the demand values. In this regard, [482] analyzes the correlation of indoor thermal performance and energy consumption. However, again, very few studies focus on residential users, i.e. households and apartments, especially with regard to dynamic data that depicts the ongoing use of that household.

In line with what other authors have noted in their work, the crucial factors that affect demand and that are to be taken into account when building predictive models are the meteorological conditions of the analyzed site. In essence, this correlation is not direct, but rather the temperature, wind speed and direction and irradiance have a significant impact on the use of heating and cooling devices, which are usually the largest consumers of energy in residential households without district heating and cooling. Besides, the current season of the year in moderate climates greatly determines what climatic conditions can be expected, and, therefore, the geographic properties of the analyzed site have to be taken into account since it is the location that determines how severe the seasonal variations in climatic conditions will be. As for the static data, the total floor space or heated volume are also said to be closely correlated with total consumption, but cannot be used to dynamically estimate demand with high time resolution. Here is where large volumes of IoT sensor data collected directly from homes can be of great help in increasing the precision of predictive models. Namely, indoor ambient temperature coupled with outdoor meteorological conditions with live occupancy data in real time can provide a precise short-term estimation of the consumption profile. Furthermore, if past behaviour is taken into account (in the form of previous demand curves both as an average over a larger time period in the past and the more current ones from the previous couple of days) with current day indicators (i.e. whether it is a working day or weekend/holiday), relatively precise hourly and possibly even inter-hourly profiles can be generated.

The presence of smart measuring devices in the form of smart plugs and cables which report real-time consumption per appliance in a home, or their substitution with an NILM algorithm as described in Subsection 3.1 where bad performance due to insufficient generalization is not an issue, provides the possibility of predicting demand on a per-appliance level. This approach is scarcely depicted in contemporary research articles with only a few papers like [310], [27] and [225] exploring this subject. Alternatively, the problem of demand forecasting is most often approached from an aggregated perspective, through the prediction of neighbourhood, city or state-level consumption, with data availability generally being the driving factor that ultimately decides what type of demand will be estimated. Time series from Figures 4, 5 and 6 illustrate the different dynamics of the demand signals from a single appliance, all appliances of one home and several aggregated homes. Since each of these applications usually requires different levels of prediction precision, the raw data used for these illustrations was averaged with different sample intervals (15 seconds, 60 seconds and 15 minutes) in accordance with the appropriate use case.

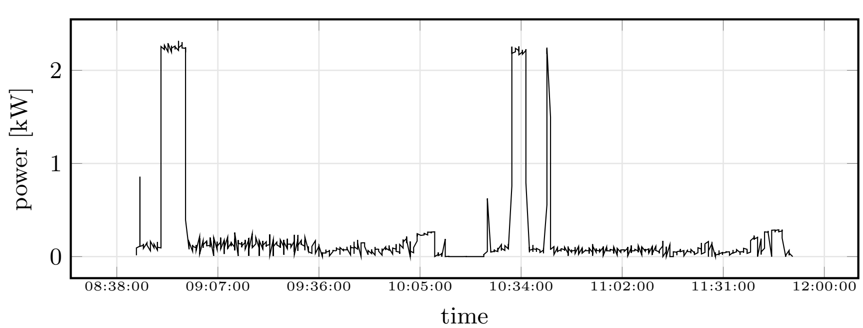

Fig. 4: Typical washing machine demand profile with 15 second averages (showing what appear to be two activations in the span of 4 hours)

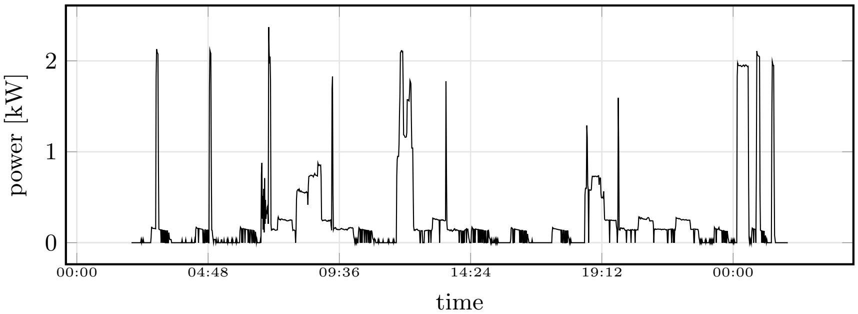

Fig. 5: Total household demand profile with 60 second averages (showing several appliance activations during a full 24-hour period)

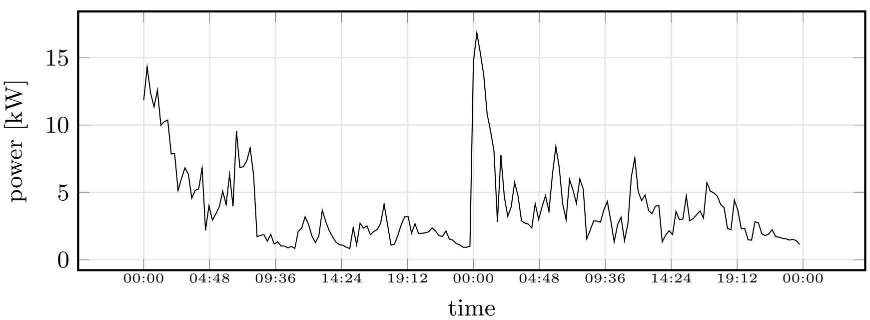

Fig. 6: Aggregate household demand profile with 15 minute averages (showing consumption for 2 days with time-of-use tariff)

4.2 Production Forecaster

It has already been mentioned that energy planning is crucial for grid stability, and that it highly depends on the forecast renewable energy sources (RES) production. Therefore, in this subsection different methodologies used for production forecasting are going to be covered as well as their relation to the field of big data.

The production of RES technologies is highly influenced by weather conditions. For example, there is very high dependency between PV production and solar radiation, similar to the relationship between wind turbines and wind speed and direction. In table 1, the selection of weather services is given followed by their main characteristics. Namely, depending on the practical application, production forecasters can have different time resolutions and horizons, which dictates the necessary weather forecast parameters. Therefore, from the abovementioned table, it can be seen that Darksky can provide estimations in terms of minutes, whilst its horizon, as some kind of compromise, is only 7 days. Additionally, depending on the approach, historical weather data might be necessary for the purpose of the training process, as, currently, the most popular approaches in the field of RES production are data-driven algorithms. Finally, the choice of weather service highly influences its price. All of those characteristics can be found in the table.

Table 1: Overview of forecasting data providers

Name	Min. forecast resolution	Max. horizon [days]	Historical data	Free up to	Coverage
OpenWeatherMap	hourly	30	Yes	60 calls/minute	Global
Weatherbit	hourly	16	Yes	500 calls/day	Global
AccuWeather	hourly	15	prev. 24h	50 calls/day	Global
Darksky	minute	7	Yes	1000 calls/day	Global
weathersteak	hourly	14	Yes	1000 calls/month	
Yahoo! Weather	hourly	10	No	2000 calls/day	Global
The Weather Channel	15 minutes	30	Yes		Global
World Weather Online	hourly	15	Yes	Not free	Global

Depending on the practical application apart from input weather parameters developed methodology varies, as well. For the use cases in which few measurements are available, physical models are usually chosen. These models are based on mathematical models and are usually deployed when there are not enough real world measurements. These models are characterized with the lowest performances in comparison with the following ones, but exist in cases of missing data. This methodology is present in the literature for various RES such as photo-voltaic panels (PVs) ([114], [332]), wind turbines (WTs) [271] and solar-thermal collectors (STCs) ([79], [392]). However, even though they do not require huge amounts of measurements, physical characteristics such as number of solar panels, position of panels and wind turbines, capacity etc. are needed and sometimes, again, inaccessible. Taking into account suppliers' tendency to equip the grid with numerous IoT sensors nowadays, the necessity of physical models is decreasing, leaving room for data-driven models, which are a more important part of this chapter and within the field of big data.

Currently the most popular and explored topic in the field of RES production forecasters is statistical and machine learning (ML) based techniques, which were proven to achieve higher performances but require substantial amounts of data. Nonetheless, bearing in mind that a huge amount of big data is currently available in the energy domain, these approaches are not common only amongst researchers but also in real practice. The first group that stands out are the statistical autoregressive methodologies SARIMA, NARIMA, ARMA, etc. [435]. They are followed by probabilistic approaches, such as in [449]. Finally, neural networks and machine learning-based approaches are proven as one of the most suitable choices ([235], [204], [450]), similar to numerous other fields.

Apart from the similar inputs regarding weather parameters and applied models for RES production forecasters, all of the methodologies are dependent on the estimation time horizon. Depending on the practical application, the orders of magnitude can range from minutes to years. Further post-processing of the obtained forecast results is another important factor. Apart from the grid control and stability, from the perspective of big data the analytical tool developed on top of the results provided by the forecaster could be exploited for failure and irregularity detection in the system together with its high level metadata. By contrast, outputs with the big time horizon could be seen as adequate for extracting conclusions on a yearly basis using big data tools already presented in this book.

4.3 Pricing Prediction

Another important application of prediction algorithms in the energy domain are price predictions. As energy sectors worldwide are becoming increasingly deregulated, variable pricing in energy trading is becoming increasingly prominent with some envisioning a not-so-distant future where the cost of energy in the wholesale and maybe even retail markets will be changing every 15 minutes while the standard nowadays is usually hourly changes at most. Having accurate predictions of wholesale market prices presents key information for large-scale energy traders because it provides an insight into future trends in the same way as stock price predictions do and allows for sound investment planning.

Wholesale price variations greatly impact retail prices, which, in turn, have a key influence on the shape of the expected demand curve from end users. Moving from fixed pricing to first time-of-use tariffs and later hourly variable pricing has allowed for energy retailers to have granular control of load levels through what is essentially implicit demand response (DR) where load increase or decrease events are defined by the current prices. Energy prices are also influenced by the availability of renewable sources. For example, systems with high PV penetration tend to have lower prices during mid-day production peaks to try and motivate users to consume more energy when there is a surplus in the system. In that way, demand predictions, production predictions and pricing productions are mutually interconnected in such a way that should result in a balanced system of equal supply and demand.

5 Conclusion

The brief overview laid out in this chapter provides an insight into some potential applications of big data-oriented tools and analytical technologies in the energy domain. With the importance of climate change mitigation growing by the day, the number of solutions working towards increasing energy efficiency and responsible energy use is only expected to rise. As such, this domain provides an interesting and challenging realm for novel research approaches.

Bibliography

1. A distributed, reliable key-value store for the most critical data of a distributed system, 2014.
2. Internet of things, 2015.
3. M. J. S. Abadi and K. Zamanifar. Producing complete modules in ontology partitioning. In *2011 International Conference on Semantic Technology and Information Retrieval*, pages 137–143. IEEE, 2011.
4. I. Abdelaziz, E. Mansour, M. Ouzzani, A. Aboulnaga, and P. Kalnis. Lusail: a system for querying linked data at scale. *Proceedings of the VLDB Endowment*, 11(4):485–498, 2017.
5. M. Acosta, O. Hartig, and J. F. Sequeda. Federated RDF query processing. In *Encyclopedia of Big Data Technologies.* Springer, Cham, 2019.
6. M. Acosta, M. Vidal, T. Lampo, J. Castillo, and E. Ruckhaus. ANAPSID: an adaptive query processing engine for SPARQL endpoints. In *The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I*, pages 18–34, 2011.
7. B. L. Agarwal. *Basic Statistics (Sixth Edition)*. New Age International, 2015.
8. M. Al Hasan and M. J. Zaki. A survey of link prediction in social networks. In *Social network data analytics*, pages 243–275. Springer, 2011.
9. M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in queensland, australia. *Advanced Engineering Informatics*, 35:1–16, 2016.
10. K. Alexander, R. Cyganiak, M. Hausenblas, and J. Zhao. Describing linked datasets. In *Proceedings of the WWW2009 Workshop on Linked Data on the Web, LDOW 2009, Madrid, Spain, April 20, 2009*, 2009.
11. H. Alharthi. Healthcare predictive analytics: An overview with a focus on saudi arabia. *Journal of Infection and Public Health*, 11(6):749–756, 2018.
12. M. Ali, M. J. Iqbal, and M. Sharif. Relationship between extreme temperature and electricity demand in pakistan. *International Journal of Energy and Environmental Engineering*, 4(1):36, 2013.
13. J. F. Allen and A. M. Frisch. What's in a semantic network? In *Proceedings of the 20th annual meeting on Association for Computational Linguistics*, pages 19–27. Association for Computational Linguistics, 1982.
14. G. Aluç, M. T. Ozsu, K. Daudjee, and O. Hartig. chameleon-db: a workload-aware robust rdf data management system. *University of Waterloo, Tech. Rep. CS-2013-10*, 2013.
15. A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, pages 290–301, 2019.
16. D. Ancin and G. Almirall. 5g to account for 15 percentage of global mobile industry by 2025 as 5g network launches accelerate, 2019.
17. G. Angeli and C. Manning. Philosophers are mortal: Inferring the truth of unseen facts. In *Proceedings of the seventeenth conference on computational natural language learning*, pages 133–142, 2013.

18. M. Arenas, P. Barceló, L. Libkin, and F. Murlak. *Foundations of Data Exchange*. Cambridge University Press, 2014.
19. M. Arenas, L. E. Bertossi, and J. Chomicki. Consistent query answers in inconsistent databases. In *PODS*, pages 68–79. ACM Press, 1999.
20. M. Arenas, J. Pérez, J. L. Reutter, and C. Riveros. Foundations of schema mapping management. In *PODS*, pages 227–238. ACM, 2010.
21. M. Arenas, J. Pérez, and E. Sallinger. Towards general representability in knowledge exchange. In *AMW*, volume 1087 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.
22. S. Arming, R. Pichler, and E. Sallinger. Combined complexity of repair checking and consistent query answering. In *AMW*, volume 1189 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
23. e. a. Assunçãoa M.D. Big data computing and clouds: Trends and future directions. *J. Parallel Distrib. Comput.*, 79–80:3–15, 2015.
24. P. Atzeni, L. Bellomarini, M. Iezzi, E. Sallinger, and A. Vlad. Weaving enterprise knowledge graphs: The case of company ownership graphs. In *EDBT*, pages 555–566. OpenProceedings.org, 2020.
25. S. Auer, L. Bühmann, C. Dirschl, O. Erling, M. Hausenblas, R. Isele, J. Lehmann, M. Martin, P. N. Mendes, B. van Nuffelen, C. Stadler, S. Tramp, and H. Williams. Managing the life-cycle of linked data with the lod2 stack. In P. Cudré-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. X. Parreira, J. Hendler, G. Schreiber, A. Bernstein, and E. Blomqvist, editors, *The Semantic Web – ISWC 2012*, pages 1–16, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
26. S. Auer, S. Scerri, A. Versteden, E. Pauwels, A. Charalambidis, S. Konstantopoulos, J. Lehmann, H. Jabeen, I. Ermilov, G. Sejdiu, A. Ikonomopoulos, S. Andronopoulos, M. Vlachogiannis, C. Pappas, A. Davettas, I. A. Klampanos, E. Grigoropoulos, V. Karkaletsis, V. de Boer, R. Siebes, M. N. Mami, S. Albani, M. Lazzarini, P. Nunes, E. Angiuli, N. Pittaras, G. Giannakopoulos, G. Argyriou, G. Stamoulis, G. Papadakis, M. Koubarakis, P. Karampiperis, A. N. Ngomo, and M. Vidal. The bigdataeurope platform - supporting the variety dimension of big data. In *Web Engineering - 17th International Conference, ICWE 2017, Rome, Italy, June 5-8, 2017, Proceedings*, pages 41–59, 2017.
27. A. Barbato, A. Capone, M. Rodolfi, and D. Tagliaferri. Forecasting the usage of household appliances through power meter sensors for demand management in the smart grid. In *2011 IEEE International Conference on Smart Grid Communications, SmartGridComm 2011*, 2011.
28. A. Bardi and P. Manghi. Enhanced publications: Data models and information systems. *Liber Quarterly*, 23(4), 2014.
29. K. S. Barsim and B. Yang. Toward a semi-supervised non-intrusive load monitoring system for event-based energy disaggregation. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 58–62, 2015. ISSN: null.
30. M. Batet, D. Sánchez, and A. Valls. An ontology-based measure to compute semantic similarity in biomedicine. *Journal of biomedical informatics*, 44(1):118–125, 2011.
31. L. E. Bayne. Big data in neonatal health care: Big reach, big reward? critical nursing clinics of north america. *Crit Care Nurs Clin North Am.*, pages 481–497, 2018.

32. C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM conference on embedded systems for energy-efficient buildings*, pages 80–89, 2014.
33. J. Bedi and D. Toshniwal. Deep learning framework to forecast electricity demand. *Applied Energy*, 238:1312–1326, 2019.
34. W. Beek, S. Schlobach, and F. van Harmelen. A contextualised semantics for owl: sameas. In *The Semantic Web. Latest Advances and New Domains - 13th International Conference, ESWC*, pages 405–419, 2016.
35. J. Begenau, M. Farboodi, and L. Veldkamp. Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 97:71–87, 2018.
36. R. Bellazzi. Big data and biomedical informatics: a challenging opportunity. *Yearb Med Inform*, 9(1):8–13, 2014.
37. G. G. S. E. Bellomarini, L. The vadalog system: Datalog-based reasoning for knowledge graphs. In *Proceedings of the VLDB Endowment*, volume 11, 2018.
38. L. Bellomarini, M. Benedetti, A. Gentili, R. Laurendi, D. Magnanimi, A. Muci, and E. Sallinger. COVID-19 and company knowledge graphs: Assessing golden powers and economic impact of selective lockdown via AI reasoning. *CoRR*, abs/2004.10119, 2020.
39. L. Bellomarini, D. Fakhoury, G. Gottlob, and E. Sallinger. Knowledge graphs and enterprise ai: the promise of an enabling technology. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 26–37. IEEE, 2019.
40. L. Bellomarini, R. R. Fayzrakhmanov, G. Gottlob, A. Kravchenko, E. Laurenza, Y. Nenov, S. Reissfelder, E. Sallinger, E. Sherkhonov, and L. Wu. Data science with vadalog: Bridging machine learning and reasoning. In *MEDI*, volume 11163 of *Lecture Notes in Computer Science*, pages 3–21. Springer, 2018.
41. L. Bellomarini, G. Gottlob, A. Pieris, and E. Sallinger. Swift logic for big data and knowledge graphs. In *IJCAI*, pages 2–10. ijcai.org, 2017.
42. L. Bellomarini, G. Gottlob, A. Pieris, and E. Sallinger. Swift logic for big data and knowledge graphs - overview of requirements, language, and system. In *SOFSEM*, volume 10706 of *Lecture Notes in Computer Science*, pages 3–16. Springer, 2018.
43. L. Bellomarini, G. Gottlob, A. Pieris, and E. Sallinger. The vadalog system: Swift logic for big data and enterprise knowledge graphs. In *AMW*, volume 2100 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.
44. L. Bellomarini, G. Gottlob, and E. Sallinger. Datalog-based reasoning for knowledge graphs. In *AMW*, volume 2369 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.
45. O. Ben-Kiki, C. Evans, and B. Ingerson. YAML Ain't Markup Language (YAML). Working Draft, Ghent University, December 2004.
46. L. Berbakov, N. Tomasevic, and M. Batic. Architecture and implementation of iot system for energy efficient living. In *26th Telecommunications Forum, TELFOR 2018, Belgrade, Serbia*, 2018.
47. Y. Berbers and W. Zwaenepoel, editors. *Google Dataset Search: Building a search engine for datasets in an open Web ecosystem*. ACM, 2019.
48. G. Berger, G. Gottlob, A. Pieris, and E. Sallinger. The space-efficient core of vadalog. In *PODS*, pages 270–284. ACM, 2019.
49. I. Bermudez, S. Traverso, M. Mellia, and M. Munafo. Exploring the cloud from passive measurements: The amazon aws case. In *2013 Proceedings IEEE INFOCOM*, pages 230–234. IEEE, 2013.
50. T. Berners-Lee. Linked data. https://www.w3.org/DesignIssues/LinkedData.html, 2006.

51. P. A. Bernstein and S. Melnik. Model management 2.0: manipulating richer mappings. In *SIGMOD Conference*, pages 1–12. ACM, 2007.
52. G. Blin, O. Curé, and D. C. Faye. A survey of rdf storage approaches. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 15, 2012.
53. A. Blumauer. From taxonomies over ontologies to knowledge graphs, 2014.
54. P. A. Bonatti, S. Decker, A. Polleres, and V. Presutti. Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). *Dagstuhl Reports*, 8(9):29–111, 2019.
55. R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza. Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models. *Applied Energy*, 208:1590–1607, 2017.
56. A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.
57. S. Borgwardt, I. I. Ceylan, and T. Lukasiewicz. Ontology-mediated queries for probabilistic databases. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
58. S. Borgwardt, I. I. Ceylan, and T. Lukasiewicz. Recent advances in querying probabilistic knowledge bases. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.* International Joint Conferences on Artificial Intelligence, 2018.
59. K. Borne. Top 10 big data challenges – a serious look at 10 big data v's, 2014.
60. L. Bornmann and R. Mutz. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science & Technology*, 66(11):2215–2222, 2015.
61. D. Borthakur et al. Hdfs architecture guide. *Hadoop Apache Project*, 53(1-13):2, 2008.
62. L. Bozzato, M. Homola, and L. Serafini. Context on the semantic web: Why and how. *ARCOE-12*, page 11, 2012.
63. F. Burgstaller, B. Neumayr, E. Sallinger, and M. Schrefl. Rule module inheritance with modification restrictions. In *OTM Conferences (2)*, volume 11230 of *Lecture Notes in Computer Science*, pages 404–422. Springer, 2018.
64. V. Bush and V. Bush. As we may think. *Resonance*, 5(11), 1945.
65. R. Cadene, H. Ben-Younes, M. Cord, and N. Thome. Murel: Multimodal relational reasoning for visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1989–1998, 2019.
66. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodriguez-Muro, R. Rosati, M. Ruzzi, and D. F. Savo. The mastro system for ontology-based data access. *Semantic Web*, 2011.
67. A. Cama, F. G. Montoya, J. Gómez, J. L. D. L. Cruz], and F. Manzano-Agugliaro. Integration of communication technologies in sensor networks to monitor the amazon environment. *Journal of Cleaner Production*, 59:32 – 42, 2013.
68. P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, and K. Tzoumas. Apache flink: Stream and batch processing in a single engine. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 36(4), 2015.
69. K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert. Is disaggregation the holy grail of energy efficiency? the case of electricity. *Energy Policy*, 52:213–234, 2013.
70. R. Catherine, B. Stephan, A. Géraldine, and B. Daniel. Weather data publication on the lod using sosa / ssn ontology. *Semantic Web*, 2019.

71. P. Ceravolo, A. Azzini, M. Angelini, and et al. Big data semantics. *Journal of Data Semantics*, 7:65–85, 2018.
72. S. Ceri, G. Gottlob, and L. Tanca. What you always wanted to know about datalog (and never dared to ask). *IEEE transactions on knowledge and data engineering*, 1(1):146–166, 1989.
73. I. I. Ceylan. *Query answering in probabilistic data and knowledge bases.* Gesellschaft für Informatik eV, 2018.
74. A. Charalambidis, A. Troumpoukis, and S. Konstantopoulos. Semagrow: optimizing federated SPARQL queries. In A. Polleres, T. Pellegrini, S. Hellmann, and J. X. Parreira, editors, *Proceedings of the 11th International Conference on Semantic Systems, SEMANTICS 2015, Vienna, Austria, September 15-17, 2015*, pages 121–128. ACM, 2015.
75. C. Chen, R. Chiang, and V. Storey. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4):1165–1188, 2012.
76. M. Chen, S. Mao, and Y. Liu. Big data: A survey. *MONET*, 19(2):171–209, 2014.
77. X. Chen, H. Chen, N. Zhang, and S. Zhang. Sparkrdf: elastic discreted rdf graph processing engine with distributed memory. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 1, pages 292–300. IEEE, 2015.
78. G. Cheng, T. Tran, and Y. Qu. Relin: relatedness and informativeness-based centrality for entity summarization. In *International Semantic Web Conference*, pages 114–129. Springer, 2011.
79. T. Chow. Performance analysis of photovoltaic-thermal collector by explicit dynamic model. *Solar Energy*, 75(2):143–152, 2003.
80. W. Chung. Review of building energy-use performance benchmarking methodologies. *Applied Energy*, 88(5):1470–1479, 2011.
81. J. Clearman, R. R. Fayzrakhmanov, G. Gottlob, Y. Nenov, S. Reissfelder, E. Sallinger, and E. Sherkhonov. Feature engineering and explainability with vadalog: A recommender systems application. In *Datalog*, volume 2368 of *CEUR Workshop Proceedings*, pages 39–43. CEUR-WS.org, 2019.
82. J. Clement. Worldwide digital population as of april 2020, 2020.
83. D. Collarana, M. Galkin, I. T. Ribón, M. Vidal, C. Lange, and S. Auer. MINTE: semantically integrating RDF graphs. In *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics, WIMS*, pages 22:1–22:11, 2017.
84. E. Commission. Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions: European data strategy, com(2020) 66 final. https://ec.europa.eu/digital-single-market/en/policies/75981/3489, 2020.
85. E. Commission. The rolling plan on ict standardisation. https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation, 2020.
86. M. Cox and D. Ellsworth. Application-controlled demand paging for out-of-core visualization. In *Proceedings. Visualization'97 (Cat. No. 97CB36155)*, pages 235–244. IEEE, 1997.
87. R. S. G. Cristina Rea. Exploratory machine learning studies for disruption prediction using large databases on diii-d. *Fusion Science and Technology*, 74(1-2), 2017.
88. T. Csar, M. Lackner, R. Pichler, and E. Sallinger. Computational social choice in the clouds. In *BTW (Workshops)*, volume P-266 of *LNI*, pages 163–167. GI, 2017.

89. T. Csar, M. Lackner, R. Pichler, and E. Sallinger. Winner determination in huge elections with mapreduce. In *AAAI*, pages 451–458. AAAI Press, 2017.
90. Q. e. a. Cui. Stochastic online learning for mobile edge computing: learning from changes. *IEEE Communications Magazine*, 57(3):63–69, 2019.
91. A. Cuzzocrea and T. Sellis. Semantics-aware approaches to big data engineering. *Journal of Data Semantics*, 6:55–56, 2017.
92. R. Cyganiak, C. Bizer, J. Garbers, O. Maresch, and C. Becker. The D2RQ Mapping Language. Technical report, FU Berlin, DERI, UCB, JP Morgan Chase, AGFA, HP Labs, Johannes Kepler Universität Linz, 2012.
93. R. Cyganiak, D. Wood, and M. Lanthaler. RDF 1.1 Concepts and Abstract Syntax. Recommendation, World Wide Web Consortium (W3C), 2014.
94. R. Dadwal, D. Graux, G. Sejdiu, H. Jabeen, and J. Lehmann. Clustering pipelines of large rdf poi data. In *European Semantic Web Conference*, pages 24–27. Springer, 2019.
95. N. Das, S. Rautaray, and M. Pandey. Big data analytics for medical applications. *International Journal of Modern Education and Computer Science*, 10, 02 2018.
96. S. Das, S. Sundara, and R. Cyganiak. R2RML: RDB to RDF Mapping Language. Working group recommendation, World Wide Web Consortium (W3C), 2012.
97. B. Data. A new world of opportunities. *NESSI White Paper*, pages 1–25, 2012.
98. B. Data. Principles and best practices of scalable realtime data systems. *N. Marz J. Warren. Henning*, 2014.
99. T. H. Davenport. Analytics 3.0. https://hbr.org/2013/12/analytics-30, 2013.
100. G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati. Using ontologies for semantic data integration. In *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, pages 187–202. Springer, 2018.
101. B. De Meester and A. Dimou. The Function Ontology. Unofficial Draft, Ghent University – imec – IDLab, 2016.
102. B. De Meester, P. Heyvaert, and A. Dimou. YARRRML. Unofficial draft, imec – Ghent University – IDLab, Aug. 2019.
103. B. De Meester, W. Maroy, A. Dimou, R. Verborgh, and E. Mannens. Declarative data transformations for Linked Data generation: the case of DBpedia. In *Proceedings of the 14^{th} ESWC*, LNCS. Springer, Cham, 2017.
104. A. Dedinec and A. Dedinec. Correlation of variables with electricity consumption data. In *Konjović, Z., Zdravković, M., Trajanović, M. (Eds.) ICIST 2016 Proceedings*, volume 1, pages 118–123. Eventonic, 2016.
105. W. V. der Scheer. 4 vs, 2015.
106. T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel. Convolutional 2d knowledge graph embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
107. B. Dhingra, Q. Jin, Z. Yang, W. W. Cohen, and R. Salakhutdinov. Neural models for reasoning over multiple mentions using coreference. *arXiv preprint arXiv:1804.05922*, 2018.
108. A. Dimou, S. Vahdati, A. Di Iorio, C. Lange, R. Verborgh, and E. Mannens. Challenges as enablers for high quality linked data: insights from the semantic publishing challenge. *PeerJ Computer Science*, 3:e105, 2017.
109. A. Dimou and M. Vander Sande. RDF Mapping Language (RML). Unofficial draft, Ghent University - iMinds - Multimedia Lab, Sept. 2014.
110. A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens, and R. Van de Walle. RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. In *Proceedings of the 7^{th} Workshop on Linked Data on the Web*, volume 1184 of *CEUR Workshop Proceedings*. CEUR, 2014.

111. A. Dimou, R. Verborgh, M. V. Sande, E. Mannens, and R. Van de Walle. Machine-interpretable dataset and service descriptions for heterogeneous data access and retrieval. In *Proceedings of the 11th International Conference on Semantic Systems*, New York, NY, USA, 2015. ACM.
112. H. Do and K. S. Cetin. Residential building energy consumption: a review of energy data availability, characteristics, and energy performance prediction methods. *Current Sustainable/Renewable Energy Reports*, 5(1):76–85, 2018.
113. A. Doan, A. Y. Halevy, and Z. G. Ives. *Principles of Data Integration.* Morgan Kaufmann, 2012.
114. A. Dolara, S. Leva, and G. Manzolini. Comparison of different physical models for pv power output prediction. *Solar energy*, 119:83–99, 2015.
115. X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, and W. Zhang. From data fusion to knowledge fusion. *Proceedings of the VLDB Endowment*, 7(10):881–892, 2014.
116. J. Dou, J. Qin, Z. Jin, and Z. Li. Knowledge graph based on domain ontology and natural language processing technology for chinese intangible cultural heritage. *Journal of Visual Languages & Computing*, 48:19–28, 2018.
117. M. Duerst and M. Suignard. Internationalized Resource Identifiers (IRIs). Standard track, IETF, Jan. 2005.
118. J. Duggan, A. J. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson, and S. Zdonik. The BigDAWG Polystore System. *SIGMOD Rec.*, 44(2):11–16, Aug. 2015.
119. M. Díaz, C. Martín, and B. Rubio. State-of-the-art, challenges, and open issues in the integration of internet of things and cloud computing. *Journal of Network and Computer Applications*, 67:99–117, 2016.
120. L. Ehrlinger and W. Wöß. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48, 2016.
121. H. Elazhary. Internet of things (iot), mobile cloud, cloudlet, mobile iot, iot cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions. *Journal of Network and Computer Applications*, 128:105–140, 2019.
122. R. Elshawi, S. Sakr, D. Talia, and P. Trunfio. Big data systems meet machine learning challenges: Towards big data science as a service. *Big Data Research*, 14:1–11, 2018.
123. C. K. Emani, N. Cullot, and C. Nicolle. Understandable big data: A survey. *Computer Science Review*, 17:70–81, 2015.
124. K. M. Endris, M. Galkin, I. Lytra, M. N. Mami, M. Vidal, and S. Auer. MULDER: querying the linked data web by bridging RDF molecule templates. In *Database and Expert Systems Applications 2017*, pages 3–18, 2017.
125. K. M. Endris, P. D. Rohde, M. Vidal, and S. Auer. Ontario: Federated query processing against a semantic data lake. In *Database and Expert Systems Applications - 30th International Conference, DEXA 2019, Linz, Austria, August 26-29, 2019, Proceedings, Part I*, pages 379–395, 2019.
126. J. Euzenat and P. Shvaiko. *Ontology Matching.* Springer, 2013.
127. R. Fagin, L. M. Haas, M. A. Hernández, R. J. Miller, L. Popa, and Y. Velegrakis. Clio: Schema mapping creation and data exchange. In *Conceptual Modeling: Foundations and Applications*, volume 5600 of *Lecture Notes in Computer Science*, pages 198–236. Springer, 2009.
128. R. Fagin, P. G. Kolaitis, L. Popa, and W. C. Tan. Composing schema mappings: Second-order dependencies to the rescue. *ACM Trans. Database Syst.*, 30(4):994–1055, 2005.

129. D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto. The agreementmakerlight ontology matching system. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 527–541. Springer, 2013.
130. D. C. Faye, O. Cure, and G. Blin. A survey of rdf storage approaches. *ARIMA Journal*, 15:11–35, 2012.
131. R. R. Fayzrakhmanov, E. Sallinger, B. Spencer, T. Furche, and G. Gottlob. Browserless web data extraction: Challenges and opportunities. In *WWW*, pages 1095–1104. ACM, 2018.
132. I. Feinerer, R. Pichler, E. Sallinger, and V. Savenkov. On the undecidability of the equivalence of second-order tuple generating dependencies. In *AMW*, volume 749 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.
133. I. Feinerer, R. Pichler, E. Sallinger, and V. Savenkov. On the undecidability of the equivalence of second-order tuple generating dependencies. *Inf. Syst.*, 48:113–129, 2015.
134. P. Figueiras, H. Antunes, G. Guerreiro, R. Costa, and R. Jardim-Gonçalves. Visualisation and Detection of Road Traffic Events Using Complex Event Processing. In *Volume 2: Advanced Manufacturing*, ASME International Mechanical Engineering Congress and Exposition, 11 2018. V002T02A081.
135. M. Figueiredo, B. Ribeiro, and A. de Almeida. Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home. *IEEE Transactions on Instrumentation and Measurement*, 63(2):364–373, 2014.
136. A. Filip. Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research. In *2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*, page 2012, 2011.
137. G. Firican. The 10 vs of big data. https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx, 2017.
138. D. Friedman. Iv.1 - the changing face of environmental monitoring. In I. Twardowska, editor, *Solid Waste: Assessment, Monitoring and Remediation*, volume 4 of *Waste Management Series*, pages 453–464. Elsevier, 2004.
139. M. Friedman, A. Y. Levy, and T. D. Millstein. Navigational plans for data integration. In *Proceedings of the IJCAI-99 Workshop on Intelligent Information Integration, Held on July 31, 1999 in conjunction with the Sixteenth International Joint Conference on Artificial Intelligence City Conference Center, Stockholm, Sweden*, 1999.
140. T. Furche, G. Gottlob, B. Neumayr, and E. Sallinger. Data wrangling for big data: Towards a lingua franca for data wrangling. In *AMW*, volume 1644 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.
141. A. Fuxman, M. A. Hernández, C. T. H. Ho, R. J. Miller, P. Papotti, and L. Popa. Nested mappings: Schema mapping reloaded. In *VLDB*, pages 67–78. ACM, 2006.
142. V. Gadepally and J. Kepner. Big data dimensional analysis. *IEEE High Performance Extreme Computing Conference*, 2014.
143. L. A. Galárraga, C. Teflioudi, K. Hose, and F. Suchanek. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*, pages 413–422, 2013.
144. J. Gao, S. Giri, E. C. Kara, and M. Bergés. Plaid: a public dataset of high-resoultion electrical appliance measurements for load identification research: demo abstract. In *proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pages 198–199, 2014.
145. Y. Gao, J. Liang, B. Han, M. Yakout, and A. Mohamed. Building a large-scale, accurate and fresh knowledge graph. In *SigKDD*, 2018.

146. N. Garg. *Apache Kafka*. Packt Publishing Ltd, 2013.
147. D. Garlaschelli, S. Battiston, M. Castri, V. Servedio, and G. Caldarelli. The scale-free topology of market investments. *Technical Report (http://www.lps.ens.fr/˜battiston/SF_TopoShareNets.pdf)*.
148. Gart2001. Gart2001, 2001.
149. A. Gastli, Y. Charabi, R. A. Alammari, and A. M. Al-Ali. Correlation between climate data and maximum electricity demand in qatar. In *2013 7th IEEE GCC Conference and Exhibition (GCC)*, pages 565–570, 2013. ISSN: null.
150. A. Gates and D. Dai. *Programming pig: Dataflow scripting with hadoop*. " O'Reilly Media, Inc.", 2016.
151. M. Ge, H. Bangui, and B. Buhnova. Big data for internet of things: A survey. *Future Generation Computer Systems*, 87:601–614, 2018.
152. N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed. Social media big data analytics: A survey. *Computers in Human Behavior*, 101:417–428, 2019.
153. M. R. Ghazi and D. Gangodkar. Hadoop, mapreduce and hdfs: a developers perspective. *Procedia Computer Science*, 48(C):45–50, 2015.
154. F. Ghofrani, Q. He, R. M. Goverde, and X. Liu. Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part C: Emerging Technologies*, 90:226–246, 2018.
155. M. Ghorbanian, S. Hacopian Dolatabadi, and P. Siano. Big data issues in smart grids: A survey. *IEEE Systems Journal*, PP:1–12, 08 2019.
156. S. Giannini. Rdf data clustering. In *International Conference on Business Information Systems*, pages 220–231. Springer, 2013.
157. S. S. Gill, S. Tuli, M. Xu, I. Singh, K. V. Singh, D. Lindsay, S. Tuli, D. Smirnova, M. Singh, U. Jain, H. Pervaiz, B. Sehgal, S. S. Kaila, S. Misra, M. S. Aslanpour, H. Mehta, V. Stankovski, and P. Garraghan. Transformative effects of iot, blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, 8:100118, 2019.
158. M. Gohar, M. Muzammal, and A. U. Rahman. Mart tss: Defining transportation system behavior using big data analytics in smart cities. *Sustainable Cities and Society*, pages 114–119, 2018.
159. B. Golshan, A. Y. Halevy, G. A. Mihaila, and W. Tan. Data Integration: After the Teenage Years. In *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA, May 14-19, 2017*, pages 101–106, 2017.
160. O. Görlitz and S. Staab. SPLENDID: SPARQL endpoint federation exploiting VOID descriptions. In *Proceedings of the Second International Workshop on Consuming Linked Data (COLD2011), Bonn, Germany, October 23, 2011*, 2011.
161. G. Gottlob, R. Pichler, and E. Sallinger. Function symbols in tuple-generating dependencies: Expressive power and computability. In *PODS*, pages 65–77. ACM, 2015.
162. G. Gottlob, R. Pichler, and E. Sallinger. Function symbols in tuple-generating dependencies: Expressive power and computability. In *AMW*, volume 1912 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2017.
163. G. Gottlob, R. Pichler, and E. Sallinger. Function symbols in tuple-generating dependencies: Expressive power and computability. In *AMW*, volume 1912 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2017.
164. G. Gottlob, A. Pieris, and E. Sallinger. Vadalog: Recent advances and applications. In *JELIA*, volume 11468 of *Lecture Notes in Computer Science*, pages 21–37. Springer, 2019.

165. S. Grant-Muller, F. Hodgson, and N. Malleson. Enhancing energy, health and security policy by extracting, enriching and interfacing next generation data in the transport domain (a study on the use of big data in cross-sectoral policy development). In *IEEE International Congress on Big Data (BigData Congress). Honolulu, HI, USA*, 2017.
166. B. C. Grau, I. Horrocks, Y. Kazakov, and U. Sattler. A logical framework for modularity of ontologies. In *IJCAI*, volume 2007, pages 298–303, 2007.
167. B. C. Grau, I. Horrocks, Y. Kazakov, and U. Sattler. Modular reuse of ontologies: Theory and practice. *Journal of Artificial Intelligence Research*, 31:273–318, 2008.
168. D. Graux, L. Jachiet, P. Geneves, and N. Layaïda. Sparqlgx: Efficient distributed evaluation of sparql with apache spark. In *International Semantic Web Conference*, pages 80–87. Springer, 2016.
169. D. Graux, L. Jachiet, P. Geneves, and N. Layaïda. A multi-criteria experimental ranking of distributed sparql evaluators. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 693–702. IEEE, 2018.
170. D. Graux, G. Sejdiu, H. Jabeen, J. Lehmann, D. Sui, D. Muhs, and J. Pfeffer. Profiting from kitties on ethereum: Leveraging blockchain RDF data with SANSA. In *SEMANTiCS Conference*, 2018.
171. S. Greco. *Multiple Criteria Decision Analysis: State of the Art Surveys*. International Series in Operations Research & Management Science. Springer-Verlag, 2005.
172. T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.
173. P. Guagliardo, R. Pichler, and E. Sallinger. Enhancing the updatability of projective views. In *AMW*, volume 1087 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.
174. V. Gudivada, M. Irfan, E. Fathi, and R. D.L. *Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning*, chapter 5. Elsevier, 2016.
175. K. Gunaratna. *Semantics-based Summarization of Entities in Knowledge Graphs*. PhD thesis, Wright State University, 2017.
176. S. Gupta, A. Arpan Kumar Kar, A. Baabdullah, and W. Al-Khowaiter. Big data with cognitive computing: A review for the future. *International Journal of Information Management*, 42:78–89, 2018.
177. P. Haase. Hybrid Enterprise Knowledge Graphs. Technical report, metaphacts GmbH, Oct. 2019.
178. R. A. A. Habeeb, F. Nasaruddin, A. Gani, I. A. T. Hashem, E. Ahmed, and M. Imran. Real-time big data processing for anomaly detection: A survey. *International Journal of Information Management*, 45:289–307, 2019.
179. R. Hai, S. Geisler, and C. Quix. Constance: An intelligent data lake system. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 2097–2100, 2016.
180. A. Y. Halevy. Answering queries using views: A survey. *VLDB J.*, 10(4):270–294, 2001.
181. S. Harispe, D. Sánchez, S. Ranwez, S. Janaqi, and J. Montmain. A framework for unifying ontology-based semantic similarity measures: A study in the biomedical domain. *Journal of biomedical informatics*, 48:38–53, 2014.
182. R. Harris. Energy star - the power to protect the environment through energy efficiency.
183. G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891., 1992.

184. O. Hartig. Reconciliation of rdf* and property graphs. *arXiv preprint arXiv:1409.3288*, 2014.
185. O. Hartig and B. Thompson. Foundations of an alternative approach to reification in rdf. *arXiv preprint arXiv:1406.3399*, 2014.
186. A. Hasan, O. Kalıpsız, and S. Akyokuş. Predicting financial market in big data: Deep learning. In *2017 International Conference on Computer Science and Engineering (UBMK)*, pages 510–515, 2017.
187. I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan. The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47:98–115, 2015.
188. B. Haslhofer, A. Isaac, and R. Simon. Knowledge graphs in the libraries and digital humanities domain. *arXiv preprint arXiv:1803.03198*, 2018.
189. L. Havrlant and V. Kreinovich. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, 46(1):27–36, 2017.
190. Y. F. R. Z. N. He, Y. Big data analytics in mobile cellular networks. *IEEE Access*, pages 1985–1996, 2016.
191. S. Heidari, Y. Simmhan, R. N. Calheiros, and R. Buyya. Scalable graph processing frameworks: A taxonomy and open challenges. *ACM Computing Surveys (CSUR)*, 51(3):1–53, 2018.
192. S. Henriet, U. Simsekli, G. Richard, and B. Fuentes. Synthetic dataset generation for non-intrusive load monitoring in commercial buildings. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, pages 1–2, 2017.
193. P. Heracleous, P. Angkititrakul, N. Kitaoka, and K. Takeda. Unsupervised energy disaggregation using conditional random fields. In *IEEE PES Innovative Smart Grid Technologies, Europe*, pages 1–5, 2014. ISSN: 2165-4824.
194. L. Hernández, C. Baladrón, J. M. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, D. J. Cook, D. Chinarro, and J. Gómez. A study of the relationship between weather variables and electric power demand inside a smart grid/smart world framework. *Sensors (Basel, Switzerland)*, 12(9):11571–11591, 2012.
195. P. Heyvaert, B. De Meester, A. Dimou, and R. Verborgh. Declarative Rules for Linked Data Generation at your Fingertips! In *Proceedings of the 15th ESWC: Posters and Demos*, 2018.
196. T. Hoff. The architecture twitter uses to deal with 150m active users, 300k qps, a 22 mb/s firehose, and send tweets in under 5 seconds. http://highscalability.com/blog/2013/7/8/the-architecture-twitter-uses-to-deal-with-150m-active-users.html, 2013.
197. S. Hoffman. *Apache Flume: distributed log collection for Hadoop*. Packt Publishing Ltd, 2013.
198. A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, et al. Knowledge graphs. *arXiv preprint arXiv:2003.02320*, 2020.
199. W. Hoiles and V. Krishnamurthy. Nonparametric demand forecasting and detection of energy aware consumers. *IEEE Transactions on Smart Grid*, 6(2):695–704, 2015.
200. Z. Hu, P. Huang, Y. Deng, Y. Gao, and E. Xing. Entity hierarchy embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 1292–1300, 2015.

201. Z. Hu, X. Ma, Z. Liu, E. Hovy, and E. Xing. Harnessing deep neural networks with logic rules. *arXiv preprint arXiv:1603.06318*, 2016.
202. Y. Huai, A. Chauhan, A. Gates, G. Hagleitner, E. N. Hanson, O. O'Malley, J. Pandey, Y. Yuan, R. Lee, and X. Zhang. Major technical advancements in apache hive. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1235–1246, 2014.
203. W. Huang, A. Chen, Y. Hsu, H. Chang, and M. Tsai. Applying market profile theory to analyze financial big data and discover financial market trading behavior - a case study of taiwan futures market. In *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, pages 166–169, 2016.
204. Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou. Comparative study of power forecasting methods for pv stations. In *2010 International Conference on Power System Technology*, pages 1–6. IEEE, 2010.
205. T. Hubauer, S. Lamparter, P. Haase, and D. M. Herzig. Use cases of the industrial knowledge graph at siemens. In *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018.
206. J. Hult. Image of the week - 8 million landsat scenes! https://www.usgs.gov/media/videos/image-week-8-million-landsat-scenes, 2018.
207. B. Humala, A. Nambi, and V. Prsad. UniversalNILM: A semi-supervised energy disaggregation framework using general appliance models. In *e-Energy*, 2018.
208. P. Hunt, M. Konar, F. P. Junqueira, and B. Reed. Zookeeper: Wait-free coordination for internet-scale systems. *USENIX annual technical conference*, 8(9), 2010.
209. T. Ibaraki and T. Kameda. On the optimal nesting order for computing n-relational joins. *ACM Trans. Database Syst.*, 9(3):482–502, 1984.
210. K. J. Imawan, A. A timeline visualization system for road traffic big data. In *IEEE International Conference on Big Data (Big Data). Santa Clara, CA, USA*, 2015.
211. T. Imieliński and W. Lipski Jr. Incomplete information in relational databases. In *Readings in Artificial Intelligence and Databases*, pages 342–360. Elsevier, 1989.
212. R. Isele and C. Bizer. Active learning of expressive linkage rules using genetic programming. *Journal of Web Semantics*, 23:2–15, 2013.
213. R. S. Istepanian and T. A.-A. zi. m-health 2.0: New perspectives on mobile health, machine learning and big data analytics. *Methods*, pages 34–40, 2018.
214. D. J. Linkedin pumps water down to its server racks, uses an interesting spine and leaf network fabric. https://www.networkworld.com/article/3161184/linkedin-pumps-water-down-to-its-server-racks-uses-an-interesting-spine-and-leaf-network-fabric.html, 2017.
215. H. Jabeen, R. Dadwal, G. Sejdiu, and J. Lehmann. Divided we stand out! forging cohorts for numeric outlier detection in large scale knowledge graphs (conod). In *European Knowledge Acquisition Workshop*, pages 534–548. Springer, 2018.
216. A. Jacobs. The pathologies of big data. acmqueue, 2009.
217. H. V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi. Big data and its technical challenges. *Commun. ACM*, 57(7):86–94, 2014.
218. P. James. Linguistic instruments in knowledge engineering rp. van de riet and ra meersman (editors)© 1992 elsevier science publishers bv all rights reserved. 97. In *Linguistic instruments in knowledge engineering: proceedings of the 1991 Workshop on Linguistic Instruments in Knowledge Engineering, Tilburg, The Netherlands, 17-18 January 1991*, page 97. North Holland, 1992.

219. V. Janev, V. Mijovic, and S. Vranes. Using the linked data approach in european e-government systems: Example from serbia. *International journal on Semantic Web and information systems*, 14(2):27–46, 2018.
220. V. Janev and S. Vraneš. The role of knowledge management solutions in enterprise business processes. *Journal of Universal Computer Science*, 11(4):526–546, 2005.
221. V. Janev and S. Vraneš. Applicability assessment of semantic web technologies. *Information Processing & Management*, 47:507–517, 2011.
222. V. S. Janev, V. Semantic web. In *Encyclopedia of Information Systems and Technology*, volume 2. Taylor and Francis, 2015.
223. M. A. Jeusfeld, M. Jarke, and J. Mylopouos. *Metamodeling for Method Engineering*. MIT Press, 2010.
224. G. Ji, S. He, L. Xu, K. Liu, and J. Zhao. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 687–696, 2015.
225. Y. Ji, E. Buechler, and R. Rajagopal. Data-driven load modeling and forecasting of residential appliances. *arXiv:1810.03727 [stat]*, 2018.
226. R. Jia, Y. Gao, and C. J. Spanos. A fully unsupervised non-intrusive load monitoring framework. In *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 872–878, 2015.
227. E. Jiménez-Ruiz and B. Cuenca Grau. Logmap: Logic-based and scalable ontology matching. In *International Semantic Web Conference*, pages 273–288. Springer, 2011.
228. E. Jiménez-Ruiz, B. C. Grau, I. Horrocks, and R. Berlanga. Logic-based assessment of the compatibility of umls ontology sources. *Journal of biomedical semantics*, 2(1):S2, 2011.
229. E. Jiménez-Ruiz, B. C. Grau, Y. Zhou, and I. Horrocks. Large-scale interactive ontology matching: Algorithms and implementation. In *ECAI*, volume 242, pages 444–449, 2012.
230. X. Jin, B. Wah, X. Cheng, and Y. Wang. Significance and challenges of big data research. *Big Data Research*, 2(2):59–64, 2015.
231. A. C. Junior, C. Debruyne, R. Brennan, and D. O'Sullivan. Funul: A method to incorporate functions into uplift mapping languages. In *Proceedings of the 18th International Conference on Information Integration and Web-Based Applications and Services*, New York, NY, USA, 2016. ACM.
232. D. Kakadia. *Apache Mesos Essentials*. Packt Publishing Ltd, 2015.
233. A. C. Kakas, R. A. Kowalski, and F. Toni. Abductive logic programming. *Journal of logic and computation*, 2(6):719–770, 1992.
234. Z. Kaoudi and I. Manolescu. Rdf in the clouds: a survey. *The VLDB Journal—The International Journal on Very Large Data Bases*, 24(1):67–91, 2015.
235. E. G. Kardakos, M. C. Alexiadis, S. I. Vagropoulos, C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis. Application of time series and artificial neural network models in short-term forecasting of pv power generation. In *2013 48th International Universities' Power Engineering Conference (UPEC)*, pages 1–6. IEEE, 2013.
236. M. Kejriwal. What is a knowledge graph? In *Domain-Specific Knowledge Graph Construction*, pages 1–7. Springer, 2019.
237. J. Kelly and W. Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2(1):1–14, 2015.
238. J. Kelly and W. J. Knottenbelt. Neural NILM: Deep neural networks applied to energy disaggregation. In *BuildSys@SenSys*, 2015.

239. M. Khabsa and C. L. Giles. The number of scholarly documents on the public web. *PloS one*, 9(5):e93949, 2014.
240. V. Khadilkar, M. Kantarcioglu, B. Thuraisingham, and P. Castagna. Jena-hbase: A distributed, scalable and efficient rdf triple store. In *Proceedings of the 11th International Semantic Web Conference Posters & Demonstrations Track, ISWC-PD*, volume 12, pages 85–88. Citeseer, 2012.
241. A. R. Khan and H. Garcia-Molina. Attribute-based crowd entity resolution. In *International Conference on Information and Knowledge Management*, pages 549–558, 2016.
242. N. Khan, A. Naim, M. R. Hussain, Q. N. Naveed, N. Ahmad, and S. Qamar. The 51 v's of big data: Survey, technologies, characteristics, opportunities, issues and challenges. In *Proceedings of the International Conference on Omni-Layer Intelligent Systems*, pages 19–24, 2019.
243. Y. Khan, A. Zimmermann, A. Jha, V. Gadepally, M. D'Aquin, and R. Sahay. One size does not fit all: Querying web polystores. *IEEE Access*, 7:9598–9617, 01 2019.
244. H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, Proceedings, pages 747–758. Society for Industrial and Applied Mathematics, 2011.
245. T. Knap, M. Kukhar, B. Machac, P. Skoda, J. Tomes, and J. Vojt. UnifiedViews: An ETL framework for sustainable RDF data processing. In *The Semantic Web: ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, Revised Selected Papers*, pages 379–383, 2014.
246. C. A. Knoblock, P. A. Szekely, J. L. Ambite, A. Goel, S. Gupta, K. Lerman, M. Muslea, M. Taheriyan, and P. Mallick. Semi-automatically mapping structured sources into the semantic web. In *Proceedings of the 9th Extended Semantic Web Conference ESWC, May 27-31, Heraklion, Crete, Greece*, pages 375–390, 2012.
247. A. Kobusińska, C. Leung, C.-H. Hsu, R. S., and V. Chang. Emerging trends, issues and challenges in internet of things, big data and cloud computing. *Future Generation Computer Systems*, 87:416–419, 2018.
248. D. Kocev, S. Džeroski, M. D. White, G. R. Newell, and P. Griffioen. Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. *Ecological Modelling*, 220(8):1159 – 1168, 2009.
249. S. Kok and P. Domingos. Learning the structure of markov logic networks. In *Proceedings of the 22nd international conference on Machine learning*, pages 441–448, 2005.
250. P. G. Kolaitis. Reflections on schema mappings, data exchange, and metadata management. In *PODS*, pages 107–109. ACM, 2018.
251. P. G. Kolaitis, R. Pichler, E. Sallinger, and V. Savenkov. Nested dependencies: structure and reasoning. In *PODS*, pages 176–187. ACM, 2014.
252. P. G. Kolaitis, R. Pichler, E. Sallinger, and V. Savenkov. Limits of schema mappings. *Theory Comput. Syst.*, 62(4):899–940, 2018.
253. D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
254. J. Z. Kolter and T. Jaakkola. Approximate inference in additive factorial HMMs with application to energy disaggregation. In *Artificial Intelligence and Statistics*, pages 1472–1482, 2012.
255. J. Z. Kolter and M. J. Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62, 2011.

256. N. Konstantinou, E. Abel, L. Bellomarini, A. Bogatu, C. Civili, E. Irfanie, M. Koehler, L. Mazilu, E. Sallinger, A. A. A. Fernandes, G. Gottlob, J. A. Keane, and N. W. Paton. VADA: an architecture for end user informed data preparation. *J. Big Data*, 6:74, 2019.
257. N. Konstantinou, M. Koehler, E. Abel, C. Civili, B. Neumayr, E. Sallinger, A. A. A. Fernandes, G. Gottlob, J. A. Keane, L. Libkin, and N. W. Paton. The VADA architecture for cost-effective data wrangling. In *SIGMOD Conference*, pages 1599–1602. ACM, 2017.
258. C. Koo, T. Hong, M. Lee, and H. Seon Park. Development of a new energy efficiency rating system for existing residential buildings. *Energy Policy*, 68:218–231, 2014.
259. D. Koo, K. Piratla, and C. J. Matthews. Towards sustainable water supply: Schematic development of big data collection using internet of things (iot). *Procedia Engineering*, 118:489–497, 2015. Defining the future of sustainability and resilience in design, engineering and construction.
260. A. Kravchenko, R. R. Fayzrakhmanov, and E. Sallinger. Web page representations and data extraction with beryl. In *ICWE Workshops*, volume 11153 of *Lecture Notes in Computer Science*, pages 22–30. Springer, 2018.
261. J. Kreps. Questioning the lambda architecture, 2014.
262. S. Krishnan and J. L. U. Gonzalez. *Building your next big thing with google cloud platform: A guide for developers and enterprise architects.* Springer, 2015.
263. D. Krompaß, S. Baier, and V. Tresp. Type-constrained representation learning in knowledge graphs. In *International Semantic Web Conference*, pages 640–655. Springer, 2015.
264. D. O. Kubitza, M. Böckmann, and D. Graux. SemanGit: A linked dataset from git. In *International Semantic Web Conference*, pages 215–228. Springer, 2019.
265. H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
266. A. Kumar, R. Shankar, A. Choudhary, and L. S. Thakur. A big data mapreduce framework for fault diagnosis in cloud-based manufacturing. *International Journal of Production Research*, 54(23):7060–7073, 2016.
267. S. Kumar and K. K. Mohbey. A review on big data based parallel and distributed approaches of pattern mining. *Journal of King Saud University - Computer and Information Sciences*, 2019.
268. A. Kumari, S. Tanwar, S. Tyagi, N. Kumar, M. Maasberg, and K.-K. R. Choo. Multimedia big data computing and internet of things applications: A taxonomy and process model. *Journal of Network and Computer Applications*, 124:169–195, 2018.
269. D. Laney. 3d data management: Controlling data volume, velocity, and variety. Application delivery strategies, Meta Group, 2001.
270. C. Lange. Krextor - An Extensible Framework for Contributing Content Math to the Web of Data. In J. H. Davenport, W. M. Farmer, J. Urban, and F. Rabe, editors, *Intelligent Computer Mathematics*, pages 304–306. Springer Berlin Heidelberg, 2011.
271. M. Lange and U. Focken. New developments in wind energy forecasting. In *2008 IEEE power and energy society general meeting-conversion and delivery of electrical energy in the 21st century*, pages 1–8. IEEE, 2008.
272. A. Langegger and W. Wöß. XLWrap – Querying and Integrating Arbitrary Spreadsheets with SPARQL. In *The Semantic Web - ISWC 2009*, pages 359–374, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

273. C. H. Lee and H.-J. Yoon. Medical big data: promise and challenges. *Kidney research and clinical practice*, 36(1):3–11, 2017.
274. J. H. Lee, A. Yilmaz, R. Denning, and T. Aldemir. Use of dynamic event trees and deep learning for real-time emergency planning in power plant operation. *Nuclear Technology*, 205(8):1035–1042, 2019.
275. M. Lee, X. He, W.-t. Yih, J. Gao, L. Deng, and P. Smolensky. Reasoning in vector space: An exploratory study of question answering. *arXiv preprint arXiv:1511.06426*, 2015.
276. M. Lefrançois, A. Zimmermann, and N. Bakerally. A SPARQL extension for generating RDF from heterogeneous formats. In *The Semantic Web 14th International Conference, ESWC 2017, Portorož, Slovenia, Proceedings*, Portoroz, Slovenia, 2017. Springer International Publishing.
277. G. W. Leibniz. The art of discovery. *Leibniz: Selections*, page 51, 1951.
278. M. Lenzerini. Data integration: A theoretical perspective. In *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 3-5, Madison, Wisconsin, USA*, pages 233–246, 2002.
279. K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas. Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50:57–70, 2020.
280. A. Y. Levy, A. Rajaraman, and J. J. Ordille. Querying heterogeneous information sources using source descriptions. In *VLDB'96, Proceedings of 22th International Conference on Very Large Data Bases, September 3-6, 1996, Mumbai (Bombay), India*, pages 251–262, 1996.
281. J. Lewis. Microservices, 2014.
282. F. Lin and W. W. Cohen. Power iteration clustering. In *Proceedings of the 27 th International Conference on Machine Learning*. Figshare, 2010.
283. S. Liu, J. McGree, Z. Ge, and Y. Xie. 8 - big data from mobile devices. In S. Liu, J. McGree, Z. Ge, and Y. Xie, editors, *Computational and Statistical Methods for Analysing Big Data with Applications*, pages 157–186. Academic Press, San Diego, 2016.
284. Y. Liu, Q. Wang, and C. Hai-Qiang. Research on it architecture of heterogeneous big data. *Journal of Applied Science and Engineering*, 18(2):135–142, 2015.
285. U. Lösch, S. Bloehdorn, and A. Rettinger. Graph kernels for rdf data. In *Extended Semantic Web Conference*, pages 134–148. Springer, 2012.
286. M. Loukides. What is data science? the future belongs to the companies and people that turn data into products. an o'reilly radar report, 2010.
287. T. Lukasiewicz, M. V. Martinez, A. Pieris, and G. I. Simari. From classical to consistent query answering under existential rules. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
288. J. Ma and J. Jiang. Applications of fault detection and diagnosis methods in nuclear power plants: A review. *Progress in Nuclear Energy*, 53:255–266, 2011.
289. M. MacDonald and S. Livengood. Benchmarking residential energy use. *Residential Buildings: Technologies, Design, and Performance Analysis*, page 12, 2020.
290. E. Makela, E. Hyvönen, and T. Ruotsalo. How to deal with massively heterogeneous cultural heritage data: Lessons learned in culturesampo. *Semantic Web*, 2012.
291. S. Makonin, I. V. Bajic, and F. Popowich. Efficient sparse matrix processing for nonintrusive load monitoring (NILM). In *NILM Workshop 2014*, 2014.
292. S. Makonin, F. Popowich, I. V. Bajić, B. Gill, and L. Bartram. Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring. *IEEE Transactions on Smart Grid*, 7(6):2575–2585, 2016.

293. M. N. Mami, D. Graux, S. Scerri, H. Jabeen, S. Auer, and J. Lehmann. Squerall: Virtual ontology-based access to heterogeneous and large data sources. In *International Semantic Web Conference*, pages 229–245. Springer, 2019.
294. P. Manghi, M. Mikulicic, and C. Atzori. De-duplication of aggregation authority files. *International Journal of Metadata, Semantics and Ontologies*, 7(2):114–130, 2012.
295. J. Manyika. Big data: The next frontier for innovation, competition, and productivity. *The McKinsey Global Institute*, pages 1—-137, 2011.
296. E. Marchi and O. Miguel. On the structure of the teaching-learning interactive process. *International Journal of Game Theory*, 3(2):83–99, 1974.
297. B. Marr. Big data and ai: 30 amazing (and free) public data sources for 2018. https://www.forbes.com/sites/bernardmarr/2018/02/26/big-data-and-ai-30-amazing-and-free-public-data-sources-for-2018, 2018.
298. R. J. Martis, V. P. Gurupur, H. Lin, A. Islam, and S. L. Fernandes. Recent advances in big data analytics, internet of things and machine learning. *Future Generation Computer Systems*, 88:696–698, 2018.
299. L. Mathews. Just how big is amazon's aws business? https://www.geek.com/chips/just-how-big-is-amazons-aws-business-hint-its-absolutely-massive-1610221/, 2014.
300. L. Mauch and B. Yang. A new approach for supervised power disaggregation by using a deep recurrent LSTM network. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 63–67, 2015. ISSN: null.
301. J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial intelligence*, 13(1-2):27–39, 1980.
302. J. McCusker, J. Erickson, K. Chastain, S. Rashid, R. Weerawarana, and D. McGuinness. What is a knowledge graph. *Semantic Web J.(submitted)*, 2018.
303. F. Michel, L. Djimenou, C. Faron-Zucker, and J. Montagnat. Translation of Heterogeneous Databases into RDF, and Application to the Construction of a SKOS Taxonomical Reference. In *International Conference on Web Information Systems and Technologies*. Springer, 2015.
304. F. Michel, L. Djimenou, C. Faron-Zucker, and J. Montagnat. xR2RML: Relational and Non-Relational Databases toRDF Mapping Language. Rapport de recherche, Laboratoire d'Informatique, Signaux et Systèmes de Sophia-Antipolis (I3S), Oct. 2017.
305. J. Michelfeit and T. Knap. Linked data fusion in ODCleanStore. In *ISWC Posters and Demonstrations Track*, 2012.
306. C. Michels, R. R. Fayzrakhmanov, M. Ley, E. Sallinger, and R. Schenkel. Oxpath-based data acquisition for dblp. In *JCDL*, pages 319–320. IEEE Computer Society, 2017.
307. V. Mijović, N. Tomašević, V. Janev, M. Stanojević, and S. Vraneš. Emergency management in critical infrastructures: A complex-event-processing paradigm. *Journal of Systems Science and Systems Engineering*, 28:37–62, 2019.
308. R. Miller. Facebook with more than two billion users on millions of servers, running thousands of configuration changes every day involving trillions of configuration checks. https://techcrunch.com/2018/07/19/how-facebook-configures-its-millions-of-servers-every-day/, 2018.
309. M. Mohammadpoor and F. Torabi. Big data analytics in oil and gas industry: An emerging trend. *Petroleum*, 2018.
310. G. Mohi Ud Din, A. U. Mauthe, and A. K. Marnerides. Appliance-level short-term load forecasting using deep neural networks. In *2018 International Conference on*

Computing, Networking and Communications (ICNC), pages 53–57, 2018. ISSN: null.
311. M. Morsey, J. Lehmann, S. Auer, and A.-C. N. Ngomo. Dbpedia sparql benchmark–performance assessment with real queries on real data. In *International semantic web conference*, pages 454–469. Springer, 2011.
312. A. Motamedi, H. Zareipour, and W. D. Rosehart. Electricity price and demand forecasting in smart grids. *IEEE Transactions on Smart Grid*, 3(2):664–674, 2012.
313. F. E. Mundial. Big data, big impact: New possibilities for international development. *Foro Económico Mundial. Cologny, Suiza. Disponible en:¡ www3. weforum. org/docs/WEF_TC_MFS_BigDataBigIm-p act_Briefing_2012. pdf*, 2012.
314. K. Munir and M. S. Anjum. The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, 14(2):116–126, 2018.
315. K. P. Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.
316. D. Murray, L. Stankovic, and V. Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4(1):1–12, 2017.
317. D. Murray, L. Stankovic, and V. Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4(1):1–12, 2017.
318. R. Mutharaju, S. Sakr, A. Sala, and P. Hitzler. D-sparq: distributed, scalable and efficient rdf query engine. *CEUR Workshop Proceedings, 261-264*, 2013.
319. P. Mutton. Cloud wars: Alibaba becomes 2nd largest hosting company. https://news.netcraft.com/archives/2017/08/22/cloud-wars-alibaba-becomes-2nd-largest-hosting-company.html, 2017.
320. M. Nasmus Sakib Khan Shabbir, M. Zawad Ali, M. Sifatul Alam Chowdhury, and X. Liang. A probabilistic approach for peak load demand forecasting. In *2018 IEEE Canadian Conference on Electrical Computer Engineering (CCECE)*, pages 1–4, 2018. ISSN: 2576-7046.
321. M. Nayyeri, C. Xu, J. Lehmann, and H. S. Yazdi. Logicenn: A neural based knowledge graphs embedding model with logical rules. *arXiv preprint arXiv:1908.07141*, 2019.
322. A.-C. N. Ngomo. Parameter-free clustering of protein-protein interaction graphs. In *Proceedings of the Fourth International Workshop on Machine Learning in Systems Biology (MLSB)*, pages 43–46. Citeseer, 2010.
323. A.-C. N. Ngomo and F. Schumacher. Borderflow: A local graph clustering algorithm for natural language processing. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 547–558. Springer, 2009.
324. M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A review of relational machine learning for knowledge graphs. *arXiv preprint arXiv:1503.00759*, 2015.
325. M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. In *ICML*, volume 11, pages 809–816, 2011.
326. M. Nickel, V. Tresp, and H.-P. Kriegel. Factorizing yago: scalable machine learning for linked data. In *Proceedings of the 21st international conference on World Wide Web*, pages 271–280. ACM, 2012.
327. T. Nikolaou, D. Kolokotsa, and G. Stavrakakis. Review on methodologies for energy benchmarking, rating and classification of buildings. *Advances in Building Energy Research*, 5(1):53–70, 2011.
328. L. Nishani and M. Biba. Statistical relational learning for collaborative filtering a state-of-the-art review. In *Natural Language Processing: Concepts, Methodologies, Tools, and Applications*, pages 688–707. IGI Global, 2020.

329. N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor. Industry-scale knowledge graphs: Lessons and challenges. *Queue*, 17(2):48–75, 2019.
330. N. F. Noy. Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record*, 33(4):65–70, 2004.
331. S. Nurdiati and C. Hoede. 25 years development of knowledge graph theory: the results and the challenge. *Memorandum*, 1876, 2008.
332. E. Ogliari, A. Dolara, G. Manzolini, and S. Leva. Physical and hybrid methods comparison for the day ahead pv output power forecast. *Renewable Energy*, 113:11–21, 2017.
333. D. Oguz, B. Ergenc, S. Yin, O. Dikenelli, and A. Hameurlain. Federated query processing on linked data: a qualitative survey and open challenges. *Knowledge Eng. Review*, 30(5):545–563, 2015.
334. A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih. Big data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4):431–448, 2018.
335. S. Owen and S. Owen. *Mahout in action*. Manning Shelter Island, NY, 2012.
336. M. T. Özsu and P. Valduriez. *Principles of Distributed Database Systems, Second Edition*. Prentice-Hall, 1999.
337. V. Palanisamy and R. Thirunavukarasu. Implications of big data analytics in developing healthcare frameworks – a review. *Journal of King Saud University - Computer and Information Sciences*, 31(4):415–425, 2019.
338. K. Panetta. Trends emerge in the gartner hype cycle for emerging technologies, 2018. *Retrieved November*, 4:5, 2018.
339. N. Papailiou, I. Konstantinou, D. Tsoumakos, and N. Koziris. H2rdf: adaptive query processing on rdf data in the cloud. In *Proceedings of the 21st International Conference on World Wide Web*, pages 397–400, 2012.
340. N. Papailiou, D. Tsoumakos, I. Konstantinou, P. Karras, and N. Koziris. H2rdf+ an efficient data management system for big rdf graphs. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 909–912, 2014.
341. C. Parker. Unexpected challenges in large scale machine learning. In *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pages 1–6, 2012.
342. O. Parson, S. Ghosh, M. Weal, and A. Rogers. An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence*, 217:1–19, 2016.
343. A. Pashazadeh and N. J. Navimipour. Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review. *Journal of Biomedical Informatics*, 82:47–62, 2018.
344. A. Patrizio. Idc: Expect 175 zettabytes of data worldwide by 2025 (december 03, 2018). "https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html.
345. H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8 (3):489–508, 2017.
346. H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508, 2017.
347. J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, 29(3):241–288, 1986.
348. M. Pennock. Digital curation: a life-cycle approach to managing and preserving usable digital information. *Library & Archives*, 1:34–45, 2007.

349. J. Pérez, R. Pichler, E. Sallinger, and V. Savenkov. Union and intersection of schema mappings. In *AMW*, volume 866 of *CEUR Workshop Proceedings*, pages 129–141. CEUR-WS.org, 2012.
350. V. Persico, A. Pescapé, A. Picariello, and G. Sperlí. Benchmarking big data architectures for social networks data processing using public cloud platforms. *Future Generation Computer Systems*, 89:98–109, 2018.
351. A. Pfandler and E. Sallinger. Distance-bounded consistent query answering. In *IJCAI*, pages 2262–2269. AAAI Press, 2015.
352. C. C. Phan, S. Big data and monitoring the grid. *The Power Grid*, pages 253–285, 2017.
353. R. Pichler, E. Sallinger, and V. Savenkov. Relaxed notions of schema mapping equivalence revisited. *Theory Comput. Syst.*, 52(3):483–541, 2013.
354. A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. In *Journal on data semantics X*, pages 133–173. Springer, 2008.
355. R. Popping. Knowledge graphs and network text analysis. *Social Science Information*, 42(1):91–106, 2003.
356. M. E. Porter and C. Advantage. Creating and sustaining superior performance. *Competitive advantage*, 167:167–206, 1985.
357. J. Pujara. *Probabilistic models for scalable knowledge graph construction.* PhD thesis, University of Maryland, College Park, MD, USA, 2016.
358. J. Pujara, B. London, L. Getoor, and W. W. Cohen. Online inference for knowledge graph construction. In *Workshop on statistical relational AI*, 2015.
359. J. Pujara, H. Miao, L. Getoor, and W. Cohen. Knowledge graph identification. In *International Semantic Web Conference*, pages 542–557. Springer, 2013.
360. J. Pujara, H. Miao, L. Getoor, and W. W. Cohen. Ontology-aware partitioning for knowledge graph identification. In *Proceedings of the 2013 workshop on Automated knowledge base construction*, pages 19–24. ACM, 2013.
361. R. Punnoose, A. Crainiceanu, and D. Rapp. Rya: a scalable rdf triple store for the clouds. In *Proceedings of the 1st International Workshop on Cloud Intelligence*, pages 1–8, 2012.
362. I. Punčochář and J. Škach. A survey of active fault diagnosis methods. *IFAC-PapersOnLine*, 51(24):1091–1098, 2018.
363. C. Quix, R. Hai, and I. Vatov. GEMMS: A generic and extensible metadata management system for data lakes. In *28th International Conference on Advanced Information Systems Engineering (CAiSE 2016)*, pages 129–136, 2016.
364. G. Radivojević, B. Lazić, and G. Šormaz. Effects of business intelligence application in tolling system. *International Journal for Traffic and Transport Engineering*, 5(1):45–53, 2015.
365. E. Rahm and H. H. Do. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, 23:3–13, 2000.
366. K. Ravi, Y. Khandelwal, B. S. Krishna, and V. Ravi. Analytics in/for cloud-an interdependence: A review. *Journal of Network and Computer Applications*, 102:17–37, 2018.
367. P. Ray. A survey on internet of things architectures. *Journal of King Saud University - Computer and Information Sciences*, 30(3):291–319, 2018.
368. G. Reali, M. Femminella, E. Nunzi, and D. Valocchi. Genomics as a service: A joint computing and networking perspective. *Computer Networks*, 145:27–51, 2018.

369. I. T. Ribón, M. Vidal, B. Kämpgen, and Y. Sure-Vetter. GADES: A graph-based semantic similarity measure. In *SEMANTICS - 12th International Conference on Semantic Systems, Leipzig, Germany*, pages 101–104, 2016.
370. E. Rich et al. Users are individuals: individualizing user models. *International journal of man-machine studies*, 18(3):199–214, 1983.
371. M. Richardson and P. Domingos. Markov logic networks. *Machine learning*, 62(1-2):107–136, 2006.
372. M. Ristevski, B. Chen. Big data analytics in medicine and healthcare. *Journal of Integrative Bioinformatics*, 15(3):–, 2018.
373. T. Rocktäschel, S. Singh, and S. Riedel. Injecting logical background knowledge into embeddings for relation extraction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1119–1129, 2015.
374. K. Rohloff and R. E. Schantz. High-performance, massively scalable distributed systems using the mapreduce software framework: The shard triple-store. In *Programming Support Innovations for Emerging Distributed Applications*, PSI EtA '10, pages 4:1–4:5, New York, NY, USA, 2010. ACM.
375. M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, and T. Bogaard. Building event-centric knowledge graphs from news. *Journal of Web Semantics*, 37:132–151, 2016.
376. J. Russell. *Cloudera Impala.* " O'Reilly Media, Inc.", 2013.
377. M. K. Saggi and S. Jain. A survey towards an integration of big data analytics to big insights for value-creation. *Inf. Process. Manage.*, 54(5):758–790, 2018.
378. M. Saleem, Y. Khan, A. Hasnain, I. Ermilov, and A. N. Ngomo. A fine-grained evaluation of SPARQL endpoint federation systems. *Semantic Web*, 7(5):493–518, 2016.
379. T. W. B. M. Saleh, I. Social-network-sourced big data analytics. *IEEE Internet Computing*, 2013.
380. E. Sallinger. Reasoning about schema mappings. In *Data Exchange, Information, and Streams*, volume 5 of *Dagstuhl Follow-Ups*, pages 97–127. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2013.
381. E. Santos, D. Faria, C. Pesquita, and F. M. Couto. Ontology alignment repair through modularization and confidence-based heuristics. *PloS one*, 10(12):e0144807, 2015.
382. A. Schätzle, M. Przyjaciel-Zablocki, T. Berberich, and G. Lausen. S2x: graph-parallel querying of rdf with graphx. In *Biomedical Data Management and Graph Online Querying*, pages 155–168. Springer, 2015.
383. A. Schätzle, M. Przyjaciel-Zablocki, S. Skilevic, and G. Lausen. S2rdf: Rdf querying with sparql on spark. *arXiv preprint arXiv:1512.07021*, 2015.
384. C. Schiano, G. Benincasa, M. Franzese, N. D. Mura], K. Pane, M. Salvatore, and C. Napoli. Epigenetic-sensitive pathways in personalized therapy of major cardiovascular diseases. *Pharmacology and Therapeutics*, page 107514, 2020.
385. E. W. Schneider. Course modularization applied: The interface system and its implications for sequence control and data analysis. *HumRRO-PP-10-73*, 1973.
386. A. Schultz, A. Matteini, R. Isele, P. N. Mendes, C. Bizer, and C. Becker. Ldif - a framework for large-scale linked data integration. In *International World Wide Web Conference*, 2012.
387. A. Schwarte, P. Haase, K. Hose, R. Schenkel, and M. Schmidt. Fedx: Optimization techniques for federated query processing on linked data. In *International Conference on The Semantic Web*, pages 601–616, 2011.

388. S. Sebastio, K. S. Trivedi, and J. Alonso. Characterizing machines lifecycle in google data centers, 2018.
389. G. Sejdiu, I. Ermilov, J. Lehmann, and M. Nadjib-Mami. DistLODStats: Distributed Computation of RDF Dataset Statistics. In *Proceedings of 17th International Semantic Web Conference*, 2018.
390. G. Sejdiu, D. Graux, I. Khan, I. Lytra, H. Jabeen, and J. Lehmann. Towards A Scalable Semantic-based Distributed Approach for SPARQL query evaluation. In *15th International Conference on Semantic Systems (SEMANTiCS)*, 2019.
391. G. Sejdiu, A. Rula, J. Lehmann, and H. Jabeen. A Scalable Framework for Quality Assessment of RDF Datasets. In *Proceedings of 18th International Semantic Web Conference*, 2019.
392. G. Serale, F. Goia, and M. Perino. Numerical model and simulation of a solar thermal collector with slurry phase change material (pcm) as the heat transfer fluid. *Solar energy*, 134:429–444, 2016.
393. T. Shafer. The 42 v's of big data and data science, 2017.
394. B. M. Shahanas and P. B. Sivakumar. Framework for a smart water management system in the context of smart city initiatives in india. *Procedia Computer Science*, 92, 142-147.
395. P. Sharma. Wireless sensor networks for environmental monitoring. In *IEERET-2014 Conference Proceedings*, 2014.
396. A. Sheth. *Panel: Data Semantics: what, where and how?*, pages 601–610. Springer US, Boston, MA, 1997.
397. A. Sheth, S. Padhee, and A. Gyrard. Knowledge graphs and knowledge networks: The story in brief. *IEEE Internet Computing*, 23(4):67–75, 2019.
398. N. J. Shoumy, L.-M. Ang, K. P. Seng, D. Rahaman, and T. Zia. Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. *Journal of Network and Computer Applications*, 149:102447, 2020.
399. A. Singh and K. B. Sahay. Short-term demand forecasting by using ANN algorithms. In *2018 International Electrical Engineering Congress (iEECON)*, pages 1–4, 2018. ISSN: null.
400. L. Singh, A. Deshpande, W. Zhou, A. Banerjee, A. Bowers, S. A. Friedler, H. V. Jagadish, G. Karypis, Z. Obradovic, A. Vullikanti, and W. Zuo. NSF BIGDATA PI meeting - domain-specific research directions and data sets. *SIGMOD Rec.*, 47(3):32–35, 2018.
401. E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web*, 5(2):51–53, 2007.
402. U. Sivarajah, M. Kamal, Z. Irani, and V. Weerakkody. Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70:263–286, 2017.
403. U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody. Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70:263–286, 2017.
404. K. Siła-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demšar, and A. S. Fotheringham. Analysis of human mobility patterns from gps trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5):881–906, 2016.
405. J. Slepicka, C. Yin, P. A. Szekely, and C. A. Knoblock. Kr2rml: An alternative interpretation of r2rml for heterogenous sources. In *Proceedings of the 6^{th} International Workshop on Consuming Linked Data (COLD)*, 2015.

406. R. Socher, D. Chen, C. D. Manning, and A. Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pages 926–934, 2013.
407. D. Souza, R. Belian, A. C. Salgado, and P. A. Tedesco. Towards a context ontology to enhance data integration processes. In *ODBIS*, pages 49–56, 2008.
408. J. F. Sowa. *Semantic networks*, chapter 1, pages 1–25. Citeseer, 1987.
409. C. Stadler, G. Sejdiu, D. Graux, and J. Lehmann. Sparklify: A Scalable Software Component for Efficient evaluation of SPARQL queries over distributed RDF datasets. In *Proceedings of 18th International Semantic Web Conference*, 2019.
410. C. Stadler, J. Unbehauen, P. Westphal, M. A. Sherif, and J. Lehmann. Simplified RDB2RDF mapping. In *Proceedings of the 8th Workshop on Linked Data on the Web (LDOW2015), Florence, Italy*, 2015.
411. D. Stepanova, M. H. Gad-Elrab, and V. T. Ho. Rule induction and reasoning over knowledge graphs. In *Reasoning Web International Summer School*, pages 142–172. Springer, 2018.
412. M. Stonebraker and J. Robertson. Big data is' buzzword du jour;'cs academics' have the best job'. *Communications of the ACM*, 56(9):10–11, 2013.
413. D. Sui, G. Sejdiu, D. Graux, and J. Lehmann. The hubs and authorities transaction network analysis using the sansa framework. In *SEMANTiCS Conference*, 2019.
414. C. Sun, R. Gao, and H. Xi. Big data based retail recommender system of non e-commerce. In *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pages 1–7, 2014.
415. Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.
416. W. P. T. Ku and H. Choi. Iot energy management platform for microgrid. In *IEEE 7th International Conference on Power and Energy Systems (ICPES). Toronto, ON, Canada*, pages 106–110, 2017.
417. S. M. Tabatabaei, S. Dick, and W. Xu. Toward non-intrusive load monitoring via multi-label classification - IEEE journals & magazine. *IEEE Transactions on Smart Grid*, 8(1):26–40, 2016.
418. A. P. Tafti, E. Behravesh, M. Assefi, E. LaRose, J. C. Badger, J. Mayer, A. Doan, D. Page, and P. L. Peissig. bignn: An open-source big data toolkit focused on biomedical sentence classification. In *IEEE International Conference on Big Data (Big Data), Boston, MA*, pages 3888–3896, 2017.
419. W. e. a. Tang. Fog-enabled smart health: Toward cooperative and secure healthcare service provision. *IEEE Communications Magazine*, 57(3):42–48, 2019.
420. M. Tasnim, D. Collarana, D. Graux, F. Orlandi, and M. Vidal. Summarizing entity temporal evolution in knowledge graphs. In *Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 961–965, 2019.
421. D. Team. Records in DBLP. http://dblp.org/statistics/recordsindblp, 2016.
422. A. Thalhammer, N. Lasierra, and A. Rettinger. Linksum: using link analysis to summarize entity data. In *International Conference on Web Engineering*, pages 244–261. Springer, 2016.
423. A. Thalhammer and A. Rettinger. Browsing dbpedia entities with summaries. In *European Semantic Web Conference*, pages 511–515. Springer, 2014.
424. A. Thusoo, D. Borthakur, and R. Murthy. Data warehousing and analytics infrastructure at facebook. In *Proc. SIGMOD '10 Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 1013–1020. ACM, 2010.

425. X. Tian, R. Han, L. Wang, G. Lu, and J. Zhan. Latency critical big data computing in finance. *The Journal of Finance and Data Science*, 1(1):33–41, 2015.
426. K. Ting and J. J. Cecho. *Apache sqoop cookbook: Unlocking hadoop for your relational database.* " O'Reilly Media, Inc.", 2013.
427. A. I. Torre-Bastida, I. L. J. Del Ser, M. Ilardia, M. N. Bilbao, and S. Campos-Cordobés. Big data for transportation and mobility: recent advances, trends and challenges. *IET Intelligent Transport Systems*, pages 742–755, 2018.
428. T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080, 2016.
429. C. Tu, X. He, Z. Shuai, and F. Jiang. Big data issues in smart grid – a review. *Renewable and Sustainable Energy Reviews*, 79:1099–1107, 2017.
430. A. M. Turing. Computing machinery and intelligence. In *Parsing the Turing Test*, pages 23–65. Springer, 2009.
431. J. D. Ullman. Information integration using logical views. In *Database Theory - ICDT '97, 6th International Conference, Delphi, Greece, January 8-10, 1997, Proceedings*, pages 19–40, 1997.
432. J. D. Ullman. Information integration using logical views. *Theor. Comput. Sci.*, 239(2):189–210, 2000.
433. S. R. Upadhyaya. Parallel approaches to machine learning - a comprehensive survey. *J. Parallel Distributed Comput.*, 73:284–292, 2013.
434. T. Urhan and M. J. Franklin. Xjoin: A reactively-scheduled pipelined join operator. *IEEE Data Eng. Bull.*, 23(2):27–33, 2000.
435. S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis. Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting. In *2016 IEEE International Energy Conference (ENERGYCON)*, pages 1–6, 2016.
436. S. Vahdati. *Collaborative Integration, Publishing and Analysis of Distributed Scholarly Metadata*. PhD thesis, Universitäts-und Landesbibliothek Bonn, 2019.
437. S. Vahdati, A. Dimou, C. Lange, and A. Di Iorio. Semantic publishing challenge: bootstrapping a value chain for scientific data. In *International Workshop on Semantic, Analytics, Visualization*, pages 73–89. Springer, 2016.
438. R. van de Riet and R. Meersman. Knowledge graphs. In *Linguistic Instruments in Knowledge Engineering: Proceedings of the 1991 Workshop on Linguistic Instruments in Knowledge Engineering, Tilburg, the Netherlands*, pages 17–18, 1991.
439. M. H. Van Emden and R. A. Kowalski. The semantics of predicate logic as a programming language. *Journal of the ACM (JACM)*, 23(4):733–742, 1976.
440. N. Van Oorschot and B. Van Leeuwen. Intelligent fire risk monitor based on linked open data, 2015.
441. J. Vater, L. Harscheidt, and A. Knoll. Smart manufacturing with prescriptive analytics. In *8th International Conference on Industrial Technology and Management (ICITM), Cambridge, United Kingdom*, pages 224–228, 2019.
442. V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth, et al. Apache hadoop yarn: Yet another resource negotiator. In *Proceedings of the 4th annual Symposium on Cloud Computing*, pages 1–16, 2013.
443. V. Verroios, H. Garcia-Molina, and Y. Papakonstantinou. Waldo: An adaptive human interface for crowd entity resolution. In *International Conference on Management of Data*, pages 1133–1148, 2017.

444. M. Vidal, K. M. Endris, S. Jazashoori, A. Sakor, and A. Rivas. Transforming heterogeneous data into knowledge for personalized treatments - A use case. *Datenbank-Spektrum*, 19(2):95–106, 2019.
445. M. Vidal, K. M. Endris, S. Jozashoori, F. Karim, and G. Palma. *Semantic Data Integration of Big Biomedical Data for Supporting of Personalised Medicine*, pages 25–56. Springer International Publishing, 2019.
446. P. H. d. Vries. *Representation of scientific texts in knowledge graphs*. PhD thesis, Groningen, 1989.
447. D. H. Vu, K. M. Muttaqi, A. P. Agalgaonkar, and A. Bouzerdoum. Intra-hour and hourly demand forecasting using selective order autoregressive model. In *2016 IEEE International Conference on Power System Technology (POWERCON)*, pages 1–6, 2016. ISSN: null.
448. V. Vychodil. *A new algorithm for computing formal concepts*. na, 2008.
449. C. Wan, J. Lin, Y. Song, Z. Xu, and G. Yang. Probabilistic forecasting of photovoltaic generation: An efficient statistical approach. *IEEE Transactions on Power Systems*, 32(3):2471–2472, 2016.
450. C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems*, 1(4):38–46, 2015.
451. C. Wang, X. Li, X. Zhou, A. Wang, and N. Nedjah. Soft computing in big data intelligent transportation systems. *Applied Soft Computing*, 38:1099–1108, 2016.
452. G. Wang, S. Yang, and Y. Han. Mashroom: end-user mashup programming using nested tables. In *Proceedings of the 18th international conference on World wide web*, pages 861–870. ACM, 2009.
453. L. Wang. Heterogeneous data and big data analytics. *Automatic Control and Information Sciences*, 3(1):8–15, 2017.
454. P. Wang, B. Xu, Y. Wu, and X. Zhou. Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, 58(1):1–38, 2015.
455. Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.
456. Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.
457. Q. Wang, B. Wang, and L. Guo. Knowledge base completion using embeddings and rules. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
458. X. Wang, W. Feng, W. Cai, H. Ren, C. Ding, and N. Zhou. Do residential building energy efficiency standards reduce energy consumption in china? - a data-driven method to validate the actual performance of buildings energy efficiency standards. *Energy Policy*, 131:82–98, 2016.
459. Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph and text jointly embedding. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1591–1601, 2014.
460. K. P. Warrior, M. Shrenik, and N. Soni. Short-term electrical load forecasting using predictive machine learning models. In *2016 IEEE Annual India Conference (INDICON)*, pages 1–6, 2016. ISSN: 2325-9418.
461. L. Weber, P. Minervini, J. Münchmeyer, U. Leser, and T. Rocktäschel. Nlprolog: Reasoning with weak unification for question answering in natural language. *arXiv preprint arXiv:1906.06187*, 2019.

462. Z. Wei, J. Zhao, K. Liu, Z. Qi, Z. Sun, and G. Tian. Large-scale knowledge base completion: Inferring via grounding network sampling over selected instances. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1331–1340. ACM, 2015.
463. L. Wen, K. Zhou, and S. Yang. Load demand forecasting of residential buildings using a deep learning model. *Electric Power Systems Research*, 179:106073, 2016.
464. K. G. Wilson. Grand challenges to computational science. *Future Generation Computer Systems*, 5(2):171–189, 1989. Grand Challenges to Computational Science.
465. S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt. Big data in smart farming – a review. *Agricultural Systems*, 153:69 – 80, 2017.
466. B. Woo, D. Vesset, C. W. Olofson, S. Conway, S. Feldman, and J. S. Bozman. Worldwide big data taxonomy. *IDC report*, 2011.
467. D. Woods. Big data requires a big, new architecture. https://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/.
468. W. Woods. What's in a link: foundations for semantic networks. *Representation and understanding*, pages 35–82, 1975.
469. M. Wylot, M. Hauswirth, P. Cudré-Mauroux, and S. Sakr. Rdf data storage and query processing schemes: A survey. *ACM Computing Surveys (CSUR)*, 51(4):1–36, 2018.
470. F. Xia, W. Wang, T. M. Bekele, and H. Liu. Big scholarly data: A survey. *IEEE Transactions on Big Data*, 3(1):18–35, 2017.
471. R. Xie, Z. Liu, and M. Sun. Representation learning of knowledge graphs with hierarchical types. In *IJCAI*, pages 2965–2971, 2016.
472. B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.
473. M. e. a. Yao. Artificial intelligence defined 5g radio access networks. *IEEE Communications Magazine*, 57(3):42–48, 2019.
474. I. Yaqoob, I. A. T. Hashem, A. Gani, S. Mokhtar, E. Ahmed, N. B. Anuar, and A. V. Vasilakos. Big data: From beginning to future. *International Journal of Information Management*, 36(6, Part B):1231–1247, 2016.
475. K. S. Yazti, D. Z. Mobile big data analytics: Research, practice, and opportunities. In *IEEE 15th International Conference on Mobile Data Management. Brisbane, QLD, Australia*, 2014.
476. M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, et al. Apache spark: a unified engine for big data processing. *Communications of the ACM*, 59(11):56–65, 2016.
477. A. Zaveri, A. Rula, R. Maurino, R. Pietrobon, J. Lehmann, , and S. Auer. Quality assessment for linked data: A survey. *Semantic Web– Interoperability, Usability, Applicability*, 7(1):63–93, 2016.
478. L. Zhang. *Knowledge graph theory and structural parsing.* Twente University Press, 2002.
479. S. Zhang, S. Jia, C. Ma, and Y. Wang. Impacts of public transportation fare reduction policy on urban public transport sharing rate based on big data analysis. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pages 280–284, 2018.
480. A. C. Zhou and B. He. *Big Data and Exascale Computing*, pages 184–187. Springer International Publishing, Cham, 2019.

481. C. Zhou, F. Su, and T. P. et al. Covid-19: Challenges to gis with big data. *Geography and Sustainability*, 03 2020.
482. J. Zhou, Y. Song, and G. Zhang. Correlation analysis of energy consumption and indoor long-term thermal environment of a residential building. *Energy Procedia*, 121:182–189, 2016.
483. Q. Zhu, X. Ma, and X. Li. Statistical learning for semantic parsing: A survey. *Big Data Mining and Analytics*, 2(4):217–239, 2019.
484. X. Zhu. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
485. P. Zikopoulos, C. Eaton, et al. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.
486. Z. Zohrevand, U. Glasser, H. Y. Shahir, M. A. Tayebi, and R. Costanzo. Hidden markov based anomaly detection for water supply systems. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1551–1560, 2016.
487. Y. Zou, T. Finin, and H. Chen. F-owl: An inference engine for semantic web. In *International Workshop on Formal Approaches to Agent-Based Systems*, pages 238–248. Springer, 2004.
488. B. C. S. C. V. J. B. B. Óskarsdóttir, M. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, pages 26–39, 2019.