

# Full Paper for ACM SenSys Proceedings

## Abstract

More than 70% of commercial buildings have giant sensor networks to monitor different aspects of building performance. However, most of these deployed sensor networks have little metadata describing context, which precludes writing analytics applications without familiarity of the particular buildings sensor metadata structure.

In this paper, we propose a technique which learns how to transform a buildings metadata to a common namespace by using a small number of examples from an expert. Once the transformation rules are learnt for one building, it can be applied across buildings with a similar metadata structure. We illustrate this on a testbed consisting of 60 buildings comprising more than 20,000 sensor points. We also illustrate how this common namespace can help a user write analytics applications that do not require building-specific knowledge, and also scales across different buildings.

## 1 Introduction

Buildings are sites of very large sensor deployments, typically containing up to several thousand sensors reporting physical measurement, continuously. Moreover, with the recent interest in reducing building energy consumption and increasing their efficiency, it is important to consider ways to quickly bootstrap a set of building data streams into an analytical pipeline. These analyses consist of jobs that help give deeper insight into the overall performance of the building, determine where there are opportunities for energy savings, and discover broken sensors. However, current ‘point’ naming conventions form a bottleneck in the scalability of the data integration process. A ‘point’ refers to a physical location where a sensor is taking measurements. Each build-

ing vendor uses their own naming scheme and unique variants of each scheme are implemented from building to building; variations exist even across buildings that have contracted the same vendor. This makes the integration process laborious for building experts and a non-starter for non experts. Moreover, the process is fundamentally unscalable. If we want to quickly run the job across many sites we need to explore methods quickly normalizing the data for acquisition.

Consider a simple analysis which has the ability to identify anomalous readings from a specific kind of sensor. To execute this job, the process organizes each sensor by type and location, organizes a the distribution of readings across types and locations, and identifies broken sensors where some fraction of their readings are above some threshold value on the distribution. In order to run this application the job needs to know the names of each sensor, its type, and its location. This information is typically encoded in the name of the sensor or ‘point’, as is common in the building system nomenclature. For example, the point BLDA1R435\_ART has all the information necessary for analysis in the form of concatenated codes. For example the name of the building (first 4 characters), the air handling unit identifier (the fifth character), the room number (R435), and the type ART (area room temperature) – which indicates that this are measurement produced by a temperature sensor – are all encoded in the aforementioned point name.

However, point names do not always follow the exact same structure and certainly do not follow the same convention across vendors. Because each point is named by a human, the names can vary. This makes it difficult to construct a general set of rules for name construction. However, although name conventions are inexact, they are generally similar across buildings with the *same vendor*, so the rules that describe the name in one building might also work to expand the names in another building. When automatic construction is not possible or uncertain, feedback from the building manager – the expert who can identify the meaning of more cryptic name encodings – could provide feedback about the metadata for the sensor. We make use of automatic name expansion and programming by example (PBE) techniques using input-output examples, in order to learn as much

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SenSys’13, November 11–15, 2013, Rome, Italy.  
Copyright © 2013 ACM 978-1-4503-1169-4 ...\$10.00

as possible from the set of names available and combine that with expert feedback to improve the certainty and coverage of our scheme.

In order to meaningfully deal with disparate building streams in a scalable fashion the streams should be *searchable* across various properties, such as building name, room location, and statistical trends. Moreover, we assert that wide searchability is necessary for achieving scalability. By providing a tool for searching across building streams, we minimize the deployment time for applications that allowing them to be used in *all* buildings, not just a single one. The aforementioned building management system user interface implicitly groups sensors by location in space or association with a system. This grouping is also captured in the name of the point used by the underlying communication protocol. For example, 'AHU' – air handling unit – is typically embedded in the name of every sensor that is associated with a particular air handling unit. A similar convention is used for denoting the type of data produced by the point (i.e. all points that contain 'ART' (area room temperature) in their name refer to a temperature sensor). However, these conventions vary slightly across buildings, making it difficult to simply integrate based on such tags alone. We need a way to unify and learn the basic set and structure of the tags in order to unify them.

In this paper, we propose a set of techniques which learns how to transform a building's metadata to a common namespace by using a small number of examples from an expert. Once the transformation rules are learnt for one building, it can be applied across buildings with a similar metadata structure. We also show how processes that extract statistical features and generate descriptive metadata can help unify data streams with respect to much deeper attributes. We show how our approach makes it easier to write applications across buildings by demonstrating its use by three different applications: 1) a rogue zone detector, 2) a broken sensor finder and 3) an application that identifies and ranks the most comfortable rooms. We illustrate these on a testbed consisting of nearly 60 buildings comprising more than 20,000 sense points. We also illustrate how this common namespace can help a user write analytics applications that do not require building-specific knowledge and scales across different buildings. We believe this is an important study given the recent trends in the penetration of the internet of things in our home and our technique can be used to unify that data for broad search and exploration of new applications.

## 2 Motivation

Buildings are notoriously complex from a management perspective. They consume a large fraction of the energy produced in the United States and much of is wasted [?]. There has been much work in the building science community to reduce their energy consumption and make them more efficient, but the route to broader impact is typically carried out through regu-

lations guided by the findings of studies in those communities [?]. We aim to let solutions reach buildings *directly* by making sense of the data they produce as quickly and accurately as possible. In order to achieve this at scale, we must explore ways to deal with the data produced from sensors within them and to enable broad analysis across several buildings at a time. Our study focuses on any building equipped with a network of sensors. Nearly three-quarters of commercial buildings contain a rich sensing fabric, installed as part of the building management system [?]. It is the data from these system and variants of it, that we wish to unify and make sense of in a more systematic and automated fashion.

The data parallels the complexity of buildings. Ad-hoc data management practices make it difficult for any analytical solution to be widely ported or run across building systems. For example, consider the following stream names: BLDA1R435\_\_ART, BLDA1R435\_\_ARS, BLDA1R545\_\_ART. Each name encodes contextual information in the form of concatenated character sequences. In these, the first 4 characters refer to the name of the building, the next one encodes the air handling unit association, the next four encode the room, and the last three encode the acronym for the type. Although the examples given are well structured, many variants within the same data set exist. For example, BLD\_1R435\_ARG\_ is the encoding for a different sensor in the same room as the others, but with a name that is *like* although not exactly the same structure as the others.

When dealing with a small number of points such differences are usually not a problem. Upon visual inspection, the two encodings are similar enough that the engineer can decode the meaning. However, for automatic processing or processing a large number of points across buildings, these kinds of variations makes it difficult to generalize the character-construction rule set. This means that analytical applications that use the data across a collection of buildings need to make building-specific adjustments in order to run the same analysis job for each building. For example, if we are trying to determine the fraction of uncomfortable rooms in each building on our campus, the two collections of sense points we need are the temperature sensors and the corresponding set points in each building. Because the code for temperature sensor and the code for set point is slightly different in each building, the query must be adjusted each time the job is set to run on the data from a new building.

At a high level we want to normalize the metadata across all buildings so that one query can run across all of buildings, and we want the normalization process to be automatic, so that the overhead in adding a new building to the set is small – the latter of which is important for achieving scale. Fundamentally, a tradeoff exists between the degree to which we can automate the normalization task and the level of coverage you get in the general rule construction. You can get full coverage with no automation. This is essentially the approach

that is common today. Every solution that incorporates a building’s data is manually adjusted, specifically for that particular building. You can get some degree of coverage if you use a general set of rules, increasing the automation factor and decrease the manual one. Tuning is quite arduous and in most cases, non-programmers are tuning the data ingestion script.

The problem for a large number of buildings is particularly pernicious to the goal. Although there are some similarities for certain kinds of sensor labels and metadata, searching across them yields results of varying success. Consider the simplest kind of a search, a grep scan across the metadata associated with the points across a set of buildings. If we wish to attain all the temperature sensors or set points for a particular building, without knowing these codes, we should be able to attain it by search for all points in a particular building that measure “temperature” or that have “setpoint” in their description. To demonstrate this by collecting all the metadata across our building testbed. The testbed consists of almost 60 buildings and over 20,000 sense points. We collected the names for each of the set points and any associated metadata that describes the meaning of the name. Then, we run a number of grep searches on the data and calculate the relative success of the query. Figure 1 shows the results for two queries. The ‘temperature’ grep string is “buildingA room temp” and the ‘setpoint’ grep string is “buildingA room setpoint”. For the temperature query we attain 87.5% of the actual sense points we are searching for, while for the setpoint query we attain only 56% of the setpoints in the building.

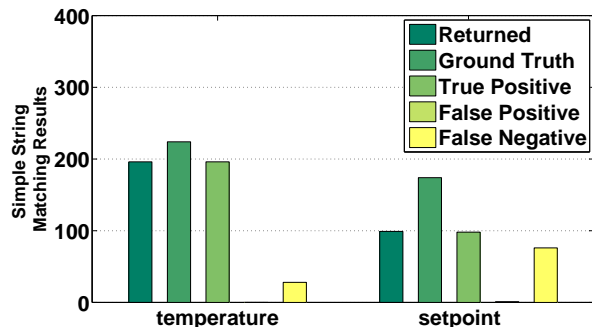


Figure 1: Results when running a grep search on the point names in a single building.

Without rule-set construction the data cannot be ingested properly or interpreted correctly. However, there are “active learning” techniques in the literature [?] that address this problem by leveraging the knowledge of an expert. The idea behind active learning is that you can request input from an expert to improve the accuracy of your algorithm. In the case of name/tag expansion you can generalize the set of rules that generate a name/tag by example, iteratively updating the name construction rule set for different types of names.

An expert feeds the system examples of names for a particular type of sensor or code, and the algorithms that generalize the character-set construction rules can raise the confidence of the expanded expression. We explore the use of *active learning* techniques to iteratively learn all the variants within an across data sets.

Fortunately, we have access to a large corpus of data from buildings on our campus. We examine the data from 56 buildings containing over 22,600 sense points. These buildings represent a vast range in age, size, and density of deployment. It also represents deployments that were set up by more than one vendor. As expected, newer buildings have many more sense points than older ones – although some old buildings that have been retrofitted have over 1000 points within them. The general trend in the number of sense points versus the year the building is built is roughly linearly increasing in the log of the number of points, as shown in Figure 2a. The maximum number of points in a single building is 6169 and the minimum is 27. The built years spans over 100 years – from 1905 to 2007. The size range spans over an order of magnitude in square footage from about 30,000 square feet to over 360,000 square feet. There is no observable correlation between the size of the building and the number of sense points. Figure 2b shows a log-log plot of the number of points versus the size of the building. There is one large building with many sense points, but the size seems to have little to do with the density of the deployment. The access this this kind of breadth of system and building types makes our study unique.

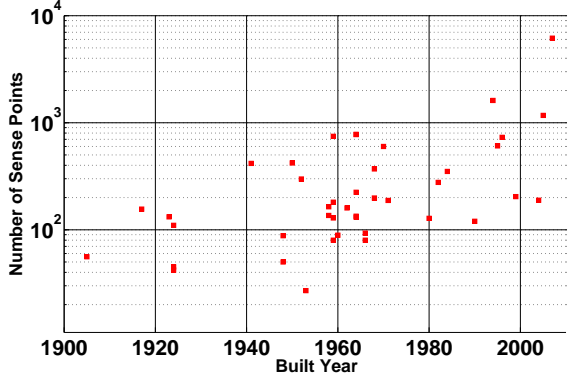
### 3 Automating Metadata Acquisition

In this section, we go into detail about how we apply program synthesis techniques and the input-output model of interaction to extract enough information from sensor names to enable sufficient coverage of sensor applications. We first provide a general overview of the technique in Section ??, followed by details of the input required, the synthesis algorithm in Section ??.

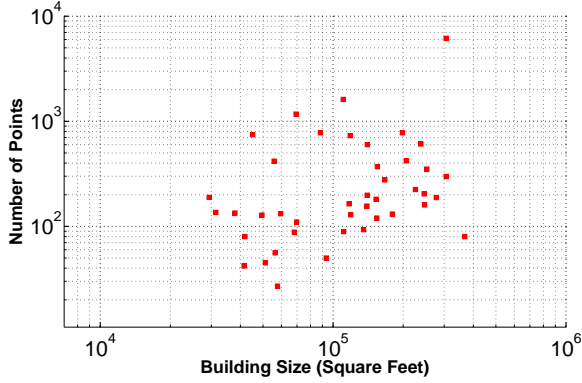
**Terminology :** We shall refer to each distinct component of a sensor name, which together signify some semantic content as a label. For instance, the example mentioned in the previous section BLD1R435\_\_ART has six labels — BLD, A, 1, R, 435 and ART. A *tag* qualifies the label to a standard namespace. In our example,

#### 3.1 Algorithm Overview

The main aim of this algorithm is to learn the regular expressions that identify different labels in a sensor name. We require an expert to give examples on how to resolve label boundaries for a particular sensor name, and the metadata tag that qualifies it. The algorithm generates a wide set of regular expressions which could extract the given label from the sensor name. When another example containing the same labels is given by the expert, the algorithm intersects the set of regular expressions obtained from the two examples, and develops and classifier which identifies which sensor names contains that particular label. As more examples are



(a) Number of Points vs Year Built



(b) Number of Points vs Building Size

Figure 2: Relationship between the built year and the building size on the number of sense points. This data is summarize from a testbed that we used for experiments that consists of almost 60 buildings and over 20,000 sense points.

added, the classifier gets tuned and so does the regular expression needed to find a label.

some intro into what we are doing to automate

### 3.2 final output - project Haystack

final output maybe in any form. we realized going into this soon that not all tags conformed to any known schema. For the general tags like room, ahu, vav we have conformed to the project haystack ones. for the remainder of the points which are building specific, we just require that a person uses a consistent schema.

Boosting the metadata can enable better usability across buildings. The question automatically becomes which metadata space to normalize to. There are many metadata schemes devised for representing all buildings. Some of them are too specific and require heavy lifting, and some of them are too simple and do not meet the criteria of being expressive enough for the necessary facets of a building.

The goal is to normalize the existing metadata.

### 3.3 formalize what we want as output from a string token

what is the input ? well the input is

- the model is we present the user an example to classify - the expert goes in and classified every tag he knows in the building. he can see the list of tags in the building while doing so. - the user also tells us which position each string corresponds to - whether a value is a constant or a variable.

### 3.4 techniques - learning from example

There is a lot of related work in obtaining structured data from unstructured data formats.

1. learning by example 2. data wrangler

- the core idea is to generate as many possible regular expressions as possible and then slowly refine our model as we get more and more examples.

### 3.5 Technique Overview

The synthesis technique is adapted from [?], and tries to learn the regular expressions which

of our technique is to provide building-specific experts the ability to come up with ur technique tried to learn the regular expression patterns that

### 3.6 learning by example

- The basic structure of our technique is derived from gluwani. -The goal is to

Our proposed technique is derived from the synthesis technique developed in [?]. In this section, we provide an overview of the technique, and then we will introduce how we adapt this technique to our problem.

At a very high level, the algorithm tries to compute all possible transformations that can transform an input string to an output string. The possibilities include a *Substring(i,j)* operation or simple position matching operation *CPos*. The two indices in the substring operation can be obtained by a regular expression, where the regular expression is

way we learn regular expressions from inputs provided by the expert. We shall use the example scada tag BLDA1R465 ART as a goto example throughout this section.

In the following sections, we will use the word *token* to refer a token from a character class. So a token might be is a character or a group of characters in a point to be expanded. In the For instance, in the point BLDA1R465 ART has the tokens as shown in Figure 3

#### 3.6.1 Inputs

For every example an expert provides, we require three types of information - (a) the normalized metadata tag which is contained in the point name, (b) the mapping of the labels in the data to normalized metadata tags , (c) the starting point of those labels, and (d) whether the value of the label is a constant or variable. For instance, consider the expert is asked to fully qualify the sensor point name BLDA1R435\_\_ART. Shown below is the expected input by the user.

The aim of the learning algorithm would then be to fully qualify the remaining examples

**Sensor Name : BLDA1R435\_\_ART**

Normalized metadata tag : <b>site</b> Position : <b>0</b>	Mapping : <b>BLD</b> Value type : <b>constant</b>
Normalized metadata tag : <b>ahu</b> Position : <b>3</b>	Mapping : <b>A</b> Value type : <b>constant</b>
Normalized metadata tag : <b>ahuRef</b> Position : <b>4</b>	Mapping : <b>1</b> Value type : <b>variable</b>
Normalized metadata tag : <b>zone</b> Position : <b>5</b>	Mapping : <b>R</b> Value type : <b>constant</b>
Normalized metadata tag : <b>zoneRef</b> Position : <b>6</b>	Mapping : <b>435</b> Value type : <b>variable</b>
Normalized metadata tag : <b>zone air temp sensor</b> Position : <b>11</b>	Mapping : <b>ART</b> Value type : <b>constant</b>

Figure 3: Input provided by the expert for fully qualifying the semantic structure of a sensor point name

Whenever the expert types in the explanation for an input, we require that the expert give a full description of the point. The full description of a point consist of the haystack tags the point contains, the starting and ending position of the string that correspond to the haystack tag, mentioning whether the substring is a constant or is variable. For instance, in our example, BLD is a constant for the haystack tag **site**, but 465 is a variable substring, because the tag value will change from point to point.

are the tokens contained in a tag, their starting and ending positions, whether the tag has an associated value, and whether the associated value

### 3.7 challenges

-different types of points all together in the same corpus, and it is not one or two . you have to generate a regex classifier for all known tags. -tokens vary from building to building -> so no pre-defined token ( alternative were using Excel's stuff or treating every letter as an individual token ). Show the experiment that shows the number of incorrectly qualified vs number of examples added. -simplest classifier to more complex classifier

### 3.8 technique we chose and modifications

learning by input output example - subtring generation -intersection -predicate generation

## 4 Evaluation of Learning By Example

In this section, we guage the effectiveness of our learning by example technique by evaluating the number of examples required to qualify labels in two large commercial buildings.

### 4.1 Test Buildings for Evaluation

We manually generated ground truth data for all the points in two buildings whose building management system was installed by different vendors. Building 1 has

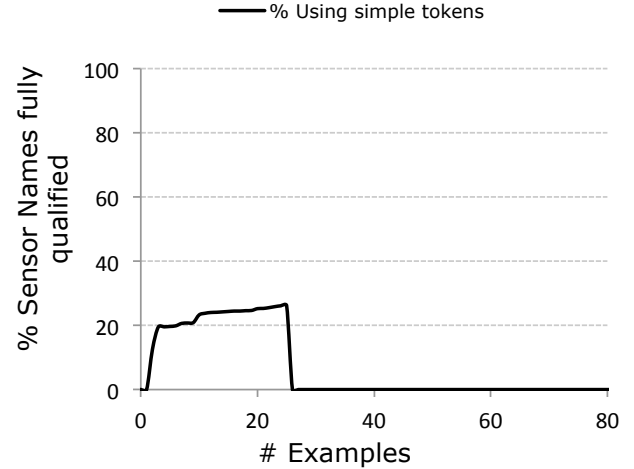


Figure 4: Number of sensor names fully qualified with a token set consisting simply of alphabetic characters [A-Z], numerals [0-9], and one token for each special character.

1586 sensor points and was built in the 1990s. Building 2 was built in the 2000s and has 2551 sense points. The label characteristics of the two buildings are shown in Figure 5.

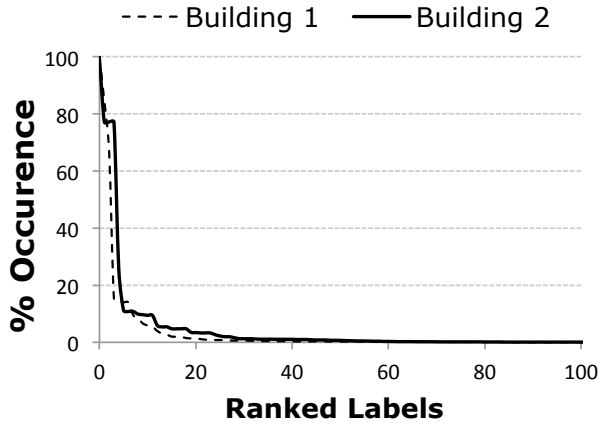
Figure 5a shows that in these two buildings, a few labels (about 20 in each building) frequently appear in a lot of sensor names. This is pretty common in commercial buildings, where a majority of the points are related to zone or room information. For instance, Building 1, has a room setpoint sensor, an airflow sensor and temperature sensor for each of its more than 200 rooms. Each of these points have a label indicating that they are a room, and a room number. These most frequent labels also fully qualify a large number of the sensor names in both buildings. In other words, learning proper classifiers and qualifications for about 20 labels could yield a full qualification for 80% of the sensor names in both these buildings.

The distribution frequency of labels also has a long tail. For both the buildings, the labels corresponding to *site* and *zoneRef* are most common. However, the distribution of labels show that there is also a long tail. These comprise of building specific sensors, alarms or status variables. There are also a lot of incomplete or inconsistent labeling of sensor names, the labels of which fall in this long tail. Thus, one of the main objectives of the learning algorithm is that it does not learn wrong classifiers for labels based on sensor names that fall in this long tail.

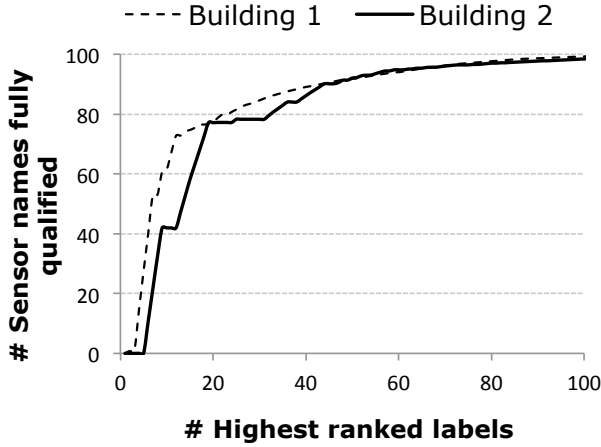
### 4.2 Convergence of labels

### 4.3 Choosing the Next Example

The large number of sensors in a building pose a challenge in selecting the next example to present to the expert. First, the expert might not always be able to browse through all sensor points to check correct qual-



(a) Percentage of sensor names each label appears in. The x-axis is sorted according to the frequency of occurrence of a label



(b) Percentage of sensor names fully qualified by the highest ranking labels. A point  $(x,y)$  indicates that  $y$  sensor names could be fully qualified by using labels ranked 1 ...  $x$

Figure 5: Characteristics of labels from two buildings we generated ground-truth data to test our learning technique

ification. Also, an expert might visally not be able to discern which points would add the most amount of information to the learning process.

This process can be facilitated by the internal notion the learning algorithm has of how much of each sensor name it has been able to qualify. This notion is maintained by simply comparing the sensor name to all the labels that the algorithm applied on it. Note that, the algorithm can incorrectly apply labels on a sensor name, which may lead to erroneous conclusions. The incorrect application of labels might be due to incorrectness or incompleteness of the boolean matching expression or the string extraction regular expressions.

We implemented four different generators to evaluate

which example should be provided next to the expert:

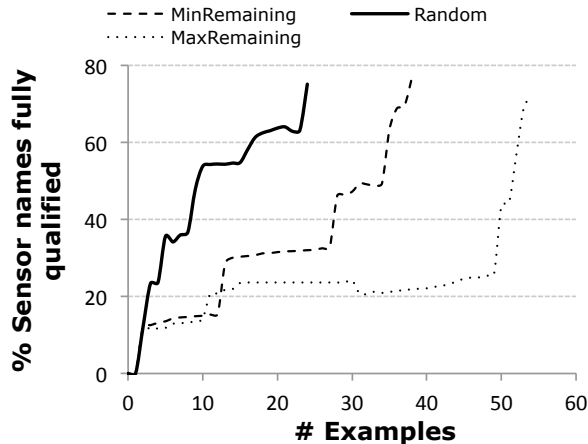
**Random:** This generator just finds at random the next example to present to the expert. While choosing the example, the random algorithm chooses among the set of sensor names which it feels it has not been able to fully qualify.

**MinRemaining :** This generator chooses the example, that according to our tool, has the minimum string length left to qualify. The intuition behind this is to gain more concrete knowledge about a small number of labels.

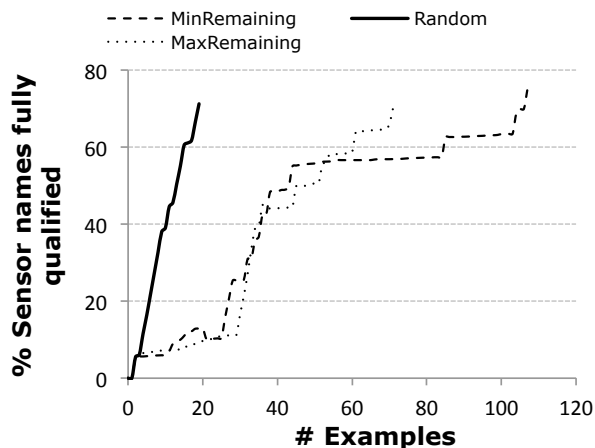
**MaxRemaining :** This chooses the example, that the learning technique feels has the maximum string length left to qualify. These examples would help the learning technique gain coverage over the space of unsees labels. The more labels the learning technique knows, the more sensor name information it will be able to qualify.

**Self-Correcting :** There are some sensor names where the learning algorithm can itself figure out that it has incorrectly qualified a sensor name. There can be three such indicators. First, for a sensor name which has matched its boolean classifier, none of its left or right position regular expressions is applicable. Second, if the sensor name has been qualified with labels that overlap over the same substring. Third, the learning algorithm can get a notion of the labelling uncertainty of a sensor name, if its qualification labels change drastically when a new example has been added. This generator gives the expert the examples that satisfy the most number of these three criteria. Once, none of the points satisfy these criteria, this generator defaults to the MinRemaining generator.

We wrote a script that automatically gave the learning algorithm the example that it asked for, and evaluated the qualifications provided by the algorithm. We terminated when the number of correct full sensor qualifications reached 70%. Figures 6 show the results of the four generators on the two buildings. The *Random* generator took the least number of examples to achieve full qualification of 70% of the sensor names, achieving it much quicker than the others. The reason for this is due to the long tail of the label distribution as was shown in Figure ???. The top 20 most occurring labels, by themselves, can fully qualify about 80% of the sensor names. A random generator has a high probability of finding one of these 80% of the points, thus acquainting itself more quickly of the most frequently occurring labels. Neither of the other three classifiers is able to achieve that. They get stuck trying to either get every bit of information from sensor with obscure labels (*MinRemaining*), or trying to cover more labels by first qualifying sensors which have been least qualified (*MaxRemaining*), or by choosing from the set of ill-formed sensor names, which would indicate errors to the learning algorithm (*Self-Correcting*).



(a) Building 1



(b) Building 2

Figure 6: The number of examples required to fully qualify 70% of sensor names in two buildings

#### 4.4 applying to other buildings

### 5 Related Work

[3] [1] [4] [2]

### 6 Case Study

In this section, we demonstrate that with the metadata automatically normalized and expanded using the techniques in previous section, we are able to implement a few applications that are generalizable from one building to another building without modification. As a proof of concept, we implement two applications: a) identify the uncomfortable rooms and b) detect rogues rooms. We also evaluate the metadata expansion technique in terms of the accuracy for both applications compared against the ground truth.

#### 6.1 Experimental Setup

We implement two applications and perform the analysis on three buildings from two campuses, and each

building is installed with a different management system. Building A and B are from the same campus on the west coast while building C is from a second campus on the east coast. Building A was built in mid-1990s using the system from Barrington [1]. Building B was recently built in 2005 using the Siemens BACnet system [2]. Building C is the newest building among the three installed with Trane’s system [3], which was put into use from 2011. We collected temperature data as well as setpoint of the rooms in each building. And the data used from each building for analysis is from one week in June 2009, January 2012 and June 2013 respectively.

#### 6.2 Uncomfortable Rooms

It’s not unusual to have rooms in a building stay extremely cold or hot thus making the occupants uncomfortable as well as incurring energy waste. The discomfort are usually caused by improper setpoint settings or dysfunction of the HVAC systems, and being able to identify these uncomfortable zones or rooms in the building is vital to occupant comfort as well as potential energy savings. With the metadata normalized using our techniques, we are able to search for the desired streams, e.g., the temperature and setpoint of a room, and analyse the thermal performance across different buildings despite of the different naming schema of the sensor or “points”.

To identify potential discomfort in a building, for each room we are particularly interested in a) how much does the temperature deviate from the comfort range? b) how much does the temperature deviate from the setpoint? c) how much does the setpoint deviate from the comfort range? To answer these questions, we first search over the points in each building for distinct temperature stream of each room and the corresponding setpoint. And then we compare the temperature with setpoint as well as the suggested comfort range from ASHRAE [4] to compute the temperature deviations in the aforementioned three different perspectives in the one-week period. We accumulate the results from all the rooms per building and generate the CDF as shown in Figure ???. Each graph illustrates how much the temperature of a building deviates from the comfort range, the setpoint and also, how much the setpoint by itself deviates from the comfort range. On average, each building is uncomfortable to some degree and we rank the rooms in each building by how much they deviates from the comfort zone and the ranking is shown in Table ???. We also present the ground truth of temperature deviation distribution in Figure 8, where we manually find temperature and setpoint streams for all the rooms to do the same analysis and generate the graphs. The ground truth analysis covers all the rooms in each building so all the potential uncomfortable rooms will be identified while the analysis using the name points expanded with our techniques would miss some of the uncomfortable rooms, since the expansion contains certain error rates.

### 6.3 Rogue Rooms

Heating and cooling contribute to the largest portion of energy consumption of a building, and often, HVAC system operates abnormally either because the system fails itself or the schedule of the building is problematic. And there are often some zones and rooms in a building that are constantly cold or hot than their neighbors and incur energy waste. We showed the temperature deviations above, and we are particularly interested in the periods when a room deviates from the setpoint more than 3 Celsius degree, which is highly likely to indicate that the room is under either heating or cooling. Therefore, for each building, we zoom in to the interested portion and find the rooms whose percentage of time that falls into the interested area is extremely high, i.e., which room's temperature deviates from the setpoint more than 3 Celsius degree in most of the time, and the results are summarized in Table ???. Again, the ground truth results are listed in Table ?? as comparison.

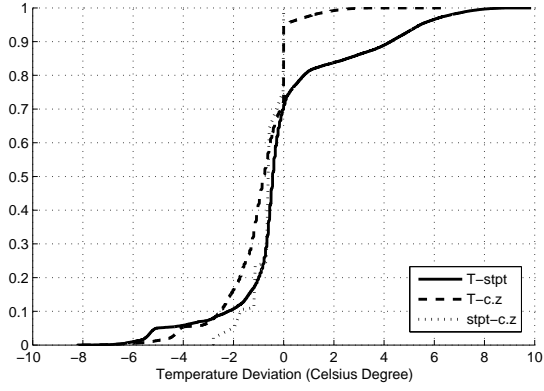
<i>BldgA</i>		<i>BldgB</i>		<i>BldgC</i>	
room#	%	room#	%	room#	%
330B	1	S3-6	1	501	1
340	1	S2-1	1	502	1
420A	1	S1-2	0.93	504	1
420	1	S7-16	0.67	557	1
698	1	S3-13	0.62	511A	1
442	0.996	S4-15	0.62	532	1
398	0.981	S5-5	0.57	224	1
336	0.96	S4-1	0.54	506	1
183	0.92	S5-16	0.48	507	1
498	0.91	S5-7	0.47	511B	1

Table 1: We present the ground truth for rooms in each building whose temperature deviates from the setpoint more than 3 Celsius degree. For each building, the first column is room number and the second column shows the percentage of the one-week time that the room deviates that much.

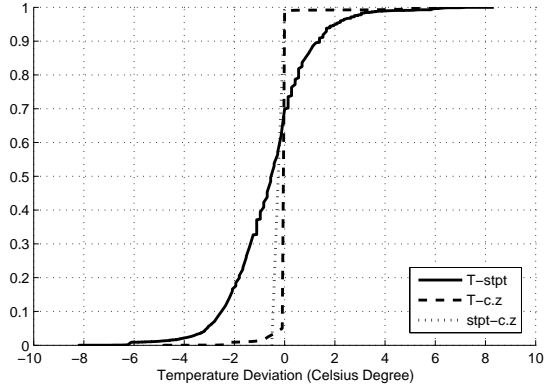
## 7 References

- [1] S. Gulwani. Automating string processing in spreadsheets using input-output examples. In *Proceedings of the 38th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL '11, pages 317–330, New York, NY, USA, 2011. ACM.
- [2] S. Gulwani, W. R. Harris, and R. Singh. Spreadsheet data manipulation using examples. In *In Communications of the ACM*, 2012.
- [3] W. R. Harris and S. Gulwani. Spreadsheet table transformations from examples. In *Proceedings of the 32Nd ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '11, pages 317–328, New York, NY, USA, 2011. ACM.
- [4] R. Singh and S. Gulwani. Learning semantic string transformations from examples. *Proc. VLDB Endow.*, 5(8):740–751, Apr. 2012.



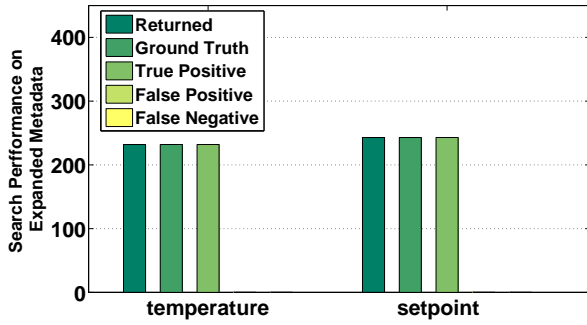


(a) Building A

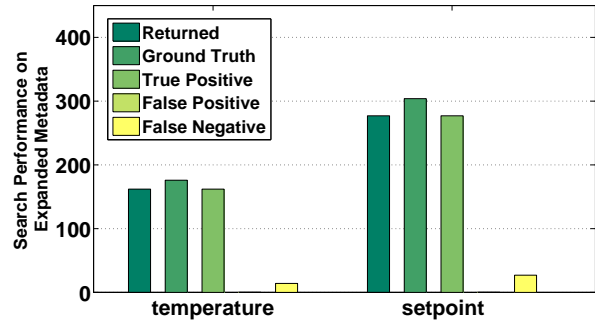


(b) Building B

Figure 7: For each building, we present the distribution of temperature deviation between: a) room temperature and the corresponding setpoint (solid), b) room temperature and the comfort range suggested by ASHRAE (dashed), c) room temperature setpoint and the ASHARE comfort suggestion (dotted).

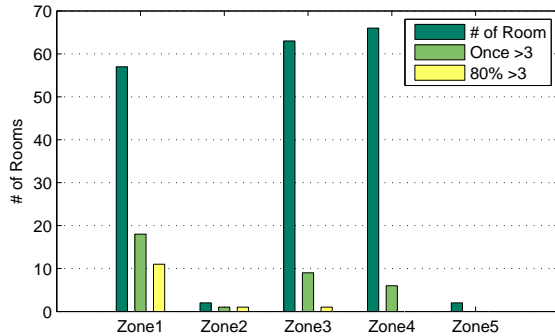


(a) Building A

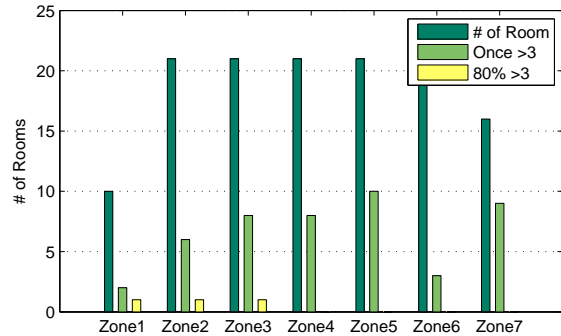


(b) Building B

Figure 8: The error rates of searches over the expanded metadata using our techniques. We do two searches particularly: “room temp” and “room temp setpoint”.



(a) Building A



(b) Building B

Figure 9: Dividing the rooms whose temperature once deviates from the setpoint more than 3 Celsius degree by HVAC zones: for each zone, we show the number of rooms in the zone, the number of zones once appears to be rogue, and the number of rooms that were rogues more than 80% in the one-week period.