**密级： 保密期限：**

xm 拷贝

**硕士学位论文**



**题目：物联网中传感器相似性搜索技术的研究**

**学 号： 2014140332**

**姓 名： 陈雪勇**

**专 业： 电子与通信工程**

**导 师： 唐碧华**

**学 院： 电子工程学院**

**年 月 日**

物联网中传感器相似性搜索技术的研究

摘 要

随着人们向物联网时代迈进，全世界有越来越多的传感器接入互联网，这些传感器的输出数据被发布到网络上，这就组成了一个即将拥有数百亿接入设备的物联网[3]。与传统互联网一样，搜索也将是物联网中的一个关键服务，能够让用户通过指定的条件去查找传感器[4]。但是在物联网中，搜索的对象是现实世界物理实体，实体状态动态变化，而传统的搜索引擎设计时主要面向的是相对静止的网络文档中的内容和元数据标签，这就决定了传统的互联网搜索方法不完全适用于物联网搜索[5]。如何让用户从这些海量的实体返回的动态实时数据中高效快速的查询到满足用户要求的实体，是当前物联网研究的重点，也是物联网发展和应用中面临的一个巨大挑战[1]。通过在物联网传感器搜索过程中设计并引用合理的排序算法，能够快速计算出传感器与查询条件的匹配程度[6]，从而有针对性的将查询重点放在评分相对更高的传感器上[14]，以此来减少通信开销，提高搜索的实时性。

已有研究成果主要是从传感器单维度的相似性进行排序，而当用户查询的实体状态需要由多个因素综合决定是，就需要引入新的排序方法。本课题提出了基于位置和内容相似性的传感器排序算法，主要面向对传感器位置和内容都有要求的搜索对象。例如当用户需要查询附近具有某一状态特征的物理实体时，此时需要综合考虑传感器的位置和内容相似性，具体可以应用在区域空气质量查询或交通拥塞查询的场景下，通过多属性的综合排序能够更加准确的查询到满足用户需求的结果。

关键字：物联网 传感器 相似性 搜索

RESEARCH AND DESIGN OF SENSOR SIMILARITY SEARCH IN THE INTERNET OF THINGS

ABSTRACT

As people move toward the Internet of Things, the world more and more sensors connected to the Internet, the output of these sensors is released to the network data, which constitutes a tens of billions of access equipment will soon have a network of things [ 3]. As with the traditional Internet, the search will also be a key service in the Internet of Things, allowing users to find the sensor through the specified conditions [4]. But in the Internet of Things, the search object is a real-world physical entity, the physical state of dynamic changes, while the traditional search engine design is mainly for relatively static network document content and metadata tags, which determines the traditional Internet search method is not fully applicable to the Internet of Things Search [5]. How to make the users retrieve the real-time data from these massive entities quickly and efficiently to the entities that meet the requirements of users is the focus of the current Internet of Things, is also a great challenge in the development and application of Internet of things [1]. By designing and referencing a reasonable sorting algorithm in the sensor network searching process, we can quickly calculate the degree of matching between the sensor and the query conditions [6], so as to focus the query on the sensor with higher scores [14]. ], In order to reduce the communication overhead, improve the real-time search.

The existing research results are mainly sorted from the similarity of the single dimension of the sensor. When the entity state of the user query needs to be determined by several factors, a new sorting method needs to be introduced. In this paper, a sorting algorithm based on position and content similarity is proposed, which is mainly aimed at the search objects that require the sensor location and content. For example, when a user needs to query a physical entity with a certain state in the vicinity, the location and content similarity of the sensor need to be taken into account. This can be applied to the regional air quality query or traffic congestion query. Sorting can be more accurate query to meet the needs of users of the results.

KEY WORDS:IOT sensor similarity search

目 录

[第一章 绪论 6](#_Toc463006009)

[2.1. 选题背景及意义 6](#_Toc463006010)

[2.2. 国内外研究动态 6](#_Toc463006011)

[2.3. 传感器相似性搜索内容及难点 6](#_Toc463006012)

[2.4. 本文主要工作及内容安排 6](#_Toc463006013)

[第二章 物联网体系架构及传感器搜索 7](#_Toc463006014)

[2.1. 物联网概念及相关技术 7](#_Toc463006015)

[2.2. 物联网三层体系架构 7](#_Toc463006016)

[2.3. 传感器搜索相关技术 7](#_Toc463006017)

[2.4. 预测分析相关技术 7](#_Toc463006018)

[2.5. 本章小结 7](#_Toc463006019)

[第三章 物联网传感器相似性搜索模型与仿真 8](#_Toc463006020)

[3.1. 传感器相似性搜索描述 8](#_Toc463006021)

[3.2. 相似性搜索模型 8](#_Toc463006022)

[3.3. 基于多属性的相似性搜索模型 8](#_Toc463006023)

[3.4. 本章小结 8](#_Toc463006024)

[第四章 基于多维的物联网传感器相似性搜索排序算法 9](#_Toc463006025)

[4.1. 传感器相似性排序算法 9](#_Toc463006026)

[4.2. 基于多维的传感器相似性排序算法 9](#_Toc463006027)

[4.3. 本章小结 9](#_Toc463006028)

[第五张 物联网搜索平台搭建及搜索验证 10](#_Toc463006029)

[5.1. 物联网搜索平台框架 10](#_Toc463006030)

[5.2. 物联网搜索平台后台搭建 10](#_Toc463006031)

[5.3. 物联网搜索平台前端搭建 10](#_Toc463006032)

[5.4. 物联网传感器相似性搜索验证 10](#_Toc463006033)

[5.5. 本章小结 10](#_Toc463006034)

[第五章 结论与展望 11](#_Toc463006035)

[参 考 文 献 11](#_Toc463006036)

## 绪论

### 选题背景及意义

随着人们向物联网时代迈进，全世界有越来越多的传感器接入互联网，这些传感器的输出数据被发布到网络上，这就组成了一个即将拥有数百亿接入设备的物联网[3]。与传统互联网一样，搜索也将是物联网中的一个关键服务，能够让用户通过指定的条件去查找传感器[4]。但是在物联网中，搜索的对象是现实世界物理实体，实体状态动态变化，而传统的搜索引擎设计时主要面向的是相对静止的网络文档中的内容和元数据标签，这就决定了传统的互联网搜索方法不完全适用于物联网搜索[5]。如何让用户从这些海量的实体返回的动态实时数据中高效快速的查询到满足用户要求的实体，是当前物联网研究的重点，也是物联网发展和应用中面临的一个巨大挑战[1]。通过在物联网传感器搜索过程中设计并引用合理的排序算法，能够快速计算出传感器与查询条件的匹配程度[6]，从而有针对性的将查询重点放在评分相对更高的传感器上[14]，以此来减少通信开销，提高搜索的实时性。

已有研究成果主要是从传感器单维度的相似性进行排序，而当用户查询的实体状态需要由多个因素综合决定是，就需要引入新的排序方法。本课题提出了基于位置和内容相似性的传感器排序算法，主要面向对传感器位置和内容都有要求的搜索对象。例如当用户需要查询附近具有某一状态特征的物理实体时，此时需要综合考虑传感器的位置和内容相似性，具体可以应用在区域空气质量查询或交通拥塞查询的场景下，通过多属性的综合排序能够更加准确的查询到满足用户需求的结果。

### 国内外研究动态

现有的物联网搜索中传感器排序方法的研究主要有基于关键字词频的排序、基于传感器内容相似性的排序和基于传感器的逻辑位置的排序。

基于关键字词频排序设计的搜索系统主要有Snoogle[7]，MAX[8]和Distributed Image Search (DIS)[9]，此类搜索系统的共性是每个传感器都需要携带有由关键字组成的文本描述信息，查询时通过查询语句中的关键字和传感器文本描述信息去匹配相关性。

基于传感器数据流相似性的研究主要有基于模糊集[5]和基于预测模型[4]的，其中基于模糊集的是先选取一个参考传感器，然后利用传感器的历史数据建立模糊集，通过计算其他传感器的模糊集与参考传感器的模糊集的相似性来对传感器进行排序。基于预测模型的是利用传感器的历史数据建立传感器的预测模型，然后通过计算在提交查询前一段时间内各测量值出现的概率，以此来预测下一个测量值与其出现的概率。

基于传感器逻辑位置的方法将传感器当前所处的逻辑位置作为主要排序依据[10]，具体做法是选取一个参考传感器，然后通过计算其他查询的传感器与参考传感器的距离来对传感器进行排序。

基于关键字词频的排序方法中，虽然已经有诸如TEDS这种传感器描述标准，但是大部分的描述都依赖于传感器的实际部署和使用（例如传感器的逻辑位置，与传感器相连的实体等），而这些信息不可能由传感器的生产厂商提供[3]。所以这些信息往往需要借助人工录入，但是人工录入的过程中难免会出现错误，而且不同的人对于同一个概念的描述方式也可能不一样，所以存在较多需要改进的地方[1]。而基于位置的方法由于局限性太强，因此一般需要配合其他方法使用。基于传感器内容相似性的方法则从一定程度上避免了基于关键字词频搜索存在的不足，因而更适用于物联网中的传感器搜索。

### 传感器相似性搜索内容及难点

### 本文主要工作及内容安排

## 物联网体系架构及传感器搜索

### 物联网概念及相关技术

### 物联网三层体系架构

### 传感器搜索相关技术

### 预测分析相关技术

### 本章小结

## 物联网传感器相似性搜索模型与仿真

### 传感器相似性搜索描述

### 相似性搜索模型

### 基于多属性的相似性搜索模型

### 本章小结

## 基于多维的物联网传感器相似性搜索排序算法

### 传感器相似性排序算法

### 基于多维的传感器相似性排序算法

### 本章小结

## 第五章 物联网搜索平台搭建

根据本文第三章所提出的物联网传感器相似性搜索架构，并结合已有的开发工具，本文设计的系统实现思路具体如下：

首先设计物联网数据爬虫对所需的互联网数据和物联网传感器数据进行抓取，采用ElasticSearch搜索服务器对获取的数据进行搜索管理。同时设计一个用户查询界面用来接收用户的查询条件，通过BosonNLP中文分词工具对该查询条件进行分析和语义分析等预处理，然后对查询词进行语义相似性计算，并给其赋予相应权值。然后利用提取到的关键词和条件范围在搜索服务器的索引库中进行搜索，以倒排索引的方式将相关度最大的前k个结果返回给用户。下面为搜索的具体实现过程。

### 系统开发环境及工具

本文借助了开源搜索服务器ElasticSearch来简化物联网搜索平台的开发搭建工作，充分利用了开源软件为项目开发带来的便利。对于开发语言的选择，本着用最合适的工作做最合适的工作的原则，吸取多种编程语言的优秀特性，为整个物联网搜索平台的搭建带来了许多便利。为了使系统结构更加清晰且利于维护，本文采用了经典的前后端分离架构来进行开发。由于搜索平台搭建所涉及到的大多数开源软件工具都是使用Java语言开发的，所以我们的后台采用Java语言作为搜索模块的框架搭建语言。并且借助python语言在编写爬虫工具方面的优势，所以系统后台用python编写了bbs数据的爬虫。同时由于Node.js在网络应用及I/O密集型应用方面的优势，能够提高搜索的实时性，所以前端部分主要采用Node.js进行搭建。

ElasticSearch是一个建立在全文搜索引擎Apache Lucene(TM) 基础上的开源搜索引擎。无论是在开源还是在专有领域，ElasticSearch都可以被认为是迄今为止最先进、性能好的功全搜索引擎库。它基于RESTful web接口提供了一个分布式多用户能力的全文搜索引擎。

ElasticSearch使用Lucene作为内部引擎，它不但包括了全文搜索功能，还可以完成以下工作：

* 分布式实时文件存储，并将每一个字段都编入索引，使其可以被检索到；
* 实时分析的分布式搜索引擎；
* 可以扩展到上百台服务器，处理PB级别的结构化数据。

目前ElasticSearch已经有很多成功案例，如维基百科，StackOverflow和GitHub都使用它来做搜索。

Node.js是一个Javascript运行环境(runtime)。实际上它是对Google V8引擎进行了封装。V8引擎执行Javascript的速度非常快，性能非常好。Node.js对一些特殊用例进行了优化，提供了替代的API，使得V8在非浏览器环境下运行得更好。Node.js是一个基于Chrome JavaScript运行时建立的平台，用于方便地搭建响应速度快、易于扩展的网络应用。Node.js 使用事件驱动，非阻塞I/O模型而得以轻量和高效，非常适合在分布式设备上运行数据密集型的实时应用。

### 物联网搜索平台后台搭建

#### 物联网爬虫设计及索引构建

本文抓取的数据主要有互联网数据、社交网数据和传感器数据，其中互联网数据主要来源于百度搜索引擎的搜索结果，社交网数据主要来源于微信公众文章和论坛bbs数据，传感器数据则来源于网络上实时发布的传感器数据。



### 物联网搜索平台前端搭建

### 物联网传感器相似性搜索验证

### 本章小结

## 第六章 系统测试及评价

## 结论与展望

## 参 考 文 献

[1].Perera, C., et al., Sensor Search Techniques for Sensing as a Service Architecture for the Internet of Things. IEEE Sensors Journal, 2014. 14(2): p. 406-420.

[2].Perera, C., et al., Context-aware Sensor Search, Selection and Ranking Model for Internet of Things Middleware. 2013 IEEE 14TH INTERNATIONAL CONFERENCE ON MOBILE DATA MANAGEMENT (MDM 2013), VOL 1, 2013: p. 314-322.

[3].Cuong, T., K. Roemer and K. Chen, Fuzzy-based Sensor Search in the Web of Things. PROCEEDINGS OF 2012 INTERNATIONAL CONFERENCE ON THE INTERNET OF THINGS, 2012: p. 127-134.

[4].Truong, C. and K. Roemer, Content-Based Sensor Search for the Web of Things, in IEEE Global Telecommunications Conference (Globecom). 2013. p. 2654-2660.

[5].Truong, C., K. Romer and K. Chen, Sensor similarity search in the Web of Things, in IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks. 2012: San Franccisco, California, USA. p. 6.

[6].Elahi, B.M., et al., Sensor Ranking: A Primitive for Efficient Content-based Sensor Search. 2009. p. 217-228.

[7].Wang, H., C.C. Tan and Q. Li, Snoogle: A Search Engine for Pervasive Environments. IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 2010. 21(8): p. 1188-1202.

[8].K.-K. Yap, V. Srinivasan, and M. Motani, B MAX: Human-centric search of the physical world,[ in Proc. 3rd Conf. Embedded Netw. Sensor Syst., 2005, pp. 166–179.

[9].T. Yan, D. Ganesan, and R. Manmatha,B Distributed image search in camera sensor networks,[ in Proc. 6th Conf. Embedded Netw. Sensor Syst., 2008, pp. 155–168.

[10].Mayer, S., D. Guinard and V. Trifa, Searching In a Web-based Infrastructure for Smart Things. PROCEEDINGS OF 2012 INTERNATIONAL CONFERENCE ON THE INTERNET OF THINGS, 2012: p. 119-126.

[11].Aberer, K.; Hauswirth, M.; Salehi, A., "Infrastructure for Data Processing in Large-Scale Interconnected Sensor Networks," in Mobile Data Management, 2007 International Conference on , vol., no., pp.198-205, 1-1 May 2007

[12].Grosky, W.I.; Kansal, A.; Nath, S.; Jie Liu; Feng Zhao, "SenseWeb: An Infrastructure for Shared Sensing," in MultiMedia, IEEE , vol.14, no.4, pp.8-13, Oct.-Dec. 2007 doi: 10.1109/MMUL.2007.82

[13].Frank, C., et al., The sensor internet at work: Locating everyday items using mobile phones. Pervasive and Mobile Computing, 2008. 4(3): p. 421-447.

[14].Romer, K.; Ostermaier, B.; Mattern, F.; Fahrmair, M.; Kellerer, W., "Real-Time Search for Real-World Entities: A Survey," in Proceedings of the IEEE , vol.98, no.11, pp.1887-1902, Nov. 2010

[15].Zhang, Daqiang; Yang, L.T.; Hongyu Huang, "Searching in Internet of Things: Vision and Challenges," in Parallel and Distributed Processing with Applications (ISPA), 2011 IEEE 9th International Symposium on , vol., no., pp.201-206, 26-28 May 2011

[16].Zhiming Ding; Xu Gao; Limin Guo; Qi Yang, "A Hybrid Search Engine Framework for the Internet of Things Based on Spatial-Temporal, Value-Based, and Keyword-Based Conditions," in Green Computing and Communications (GreenCom), 2012 IEEE International Conference on , vol., no., pp.17-25, 20-23 Nov. 2012

[17].Govindaraju, V.; Chen-Han Ho; Nowatzki, T.; Chhugani, J.; Satish, N.; Sankaralingam, K.; Changkyu Kim, "DySER: Unifying Functionality and Parallelism Specialization for Energy-Efficient Computing," in Micro, IEEE , vol.32, no.5, pp.38-51, Sept.-Oct. 2012

[18].Yongtao Ma; Liuji Zhou; Kaihua Liu; Jinlong Wang, "Iterative Phase Reconstruction and Weighted Localization Algorithm for Indoor RFID-Based Localization in NLOS Environment," in Sensors Journal, IEEE , vol.14, no.2, pp.597-611, Feb. 2014