Research and Implementation of LED Optical Design Information Integration and Sharing Service Platform

Jing GONG\textsuperscript{1,a,*}, Cai-Feng CAO\textsuperscript{2,b}

\textsuperscript{1,2}Computer School Wuyi University Jiangmen, Guangdong, China
\textsuperscript{a}gongjing457@163.com, \textsuperscript{b}cfcao@126.com
*Corresponding author

Keywords: LED Optical Design, Information Sharing Service Platform, NET, Focused-crawler.

Abstract. Led optical design information integration and sharing service platform, provides resource sharing services for led optical design and provides for the industry, school, research institute and users with service and communication. It integrates two times optical design information, and uses the led optical design theme crawler algorithm which is put forward by us to collect data accurately and search led topic relevant web information.

Background of the Research

According to the requirement of jiangmen economy and information bureau’s project and jiangmen led industry development plan, the study uses the technologies of internet, web, database and web crawler to establish an Led optical design information sharing service platform system. It will contribute to improve led optical design technology, shorten led product design and manufacturing cycle, reduce production costs and improve the overall technological level and market competitiveness of Jiangmen led industry.

Data Information Integration Design

In the process of led optical design, a large number of important data will be generated. The system data information planning is as follows:

The first category: material database of the first time led optical design. The library includes chip data table, scaffold material data table and packaging material data table. The chip data table includes chip type, name, specifications, performance parameters, nature, manufacturers, appearance pictures data; scaffold material data table data includes a code, specification, model, name, material type, nature, manufacturers, appearance pictures; packaging materials data table data includes packaging material number, specification, model, name, type, effect of encapsulation, manufacturers, appearance pictures.

The second category: information database of the secondary led optical design. By the first time optical design, it produces a number of necessary design results, based on the results to realize the secondary optical design and to achieve the ultimate purpose of product design. Information database of the secondary led optical design contains a plurality of intermediate results data table, the main data tables are listed below: led light source package data table, light cup data table, reflective material data table, lens data table, circuit board data table.

The third category: led optical design product information database. Secondary optical design produces led optical design products; led optical design product information database contains product number, name, product description, design documents, performance indicators, responsible person, date, production, images).
The fourth category: led optical design technology progress information database. By topic crawler search software, get the latest progress of led optical design technology information, and saved in the database. It contains information number, title, links page, source, priority, date.

**System Function Design**

The overall framework of led optical design information integration and sharing service platform is shown in Fig.1.

![Figure 1. Overall function structure framework of LED optical design information integration and sharing service platform.](image)

This system function contains six modules, contains register login module, design platform introduction module, service providing module, optical design module, led optical design progress information module, backstage management module.

**LED Optical Design Focused Crawler Algorithm Design**

To design led optical design theme crawler algorithm to achieve effective fetch and retrieval led optical design technology progress information.

*Algorithm:* the improved *shark-search* algorithm removes anchor text near the character factors, inherited his father's reservations relating to the similarity of nodes and anchor text relevance of these two factors, in order to improve the efficiency of its crawled pages. Calculation of the reliance of the sub-topic page is shown in Eq.1.

\[
\text{potential\_score}(\text{child\_url}) = \gamma \times \text{inherited\_score}(\text{child\_url}) + (1-\gamma) \times \text{sim}(q, \text{anchor}).
\]  

(1)

The coefficient of $\gamma < 1$, here $\gamma = 0.5$, $\delta = 0.5$. In Eq.1, \text{potential\_score}(child\_url) is sub-pages topic relevance. \text{inherited\_score}(child\_url) is topic relevance of sub-pages inherited from father page. \text{Inherited\_score}(child\_url) $= \delta \times \text{sim}(q, \text{current\_url})$, if \text{sim}(q, \text{current\_url}) $> \delta$,

\[\delta \times \text{inherited\_score}(\text{current\_url}), \text{other}\]  

(2)
\( \text{sim}(q, \text{anchor}) \) is comparability between anchor text and led optical design topic vector q, anchor is anchor text. \( \text{inherited \_score}(\text{child \_url}) \) and \( \text{sim}(q, \text{anchor}) \) are calculated with the same original Shark-Search algorithm\[1\].

Calculate sub-page url thematic relevance score, then compare the correlation scores with the threshold value, if the correlation score is greater than or equal to the threshold value, put sub-page url into the url to be crawled list waiting to crawl, and vice versa, then discard the sub-page url. The description of led focused crawler algorithm (topic-first) is as follows.

Input: url list of father node \( \text{fatherQueue} \);
Output: led optical design topic-relative web pages sets \( \text{led} \);
while(\( \text{fatherQueue} \) not null ) {
    get an \( \text{fatherURL} \) from \( \text{fatherQueue} \);
    crawl pages of \( \text{fatherURL} \) and extract all the urls, put urls into URL list of child node \( \text{sonQueue} \);
    while(\( \text{sonQueue} \) not null) {
        get a \( \text{sonQueue} \) url;
        //according to \( \text{Eq.2, Eq.1} \), calculate \( \text{ingerited \_score} \) and potential_score
        if(\( \text{sonURL} \) contains Topic - mark the word){
            put \( \text{sonURL} \) into \( \text{fatherQueue} \); }
        If(potential_score\( \geq \)threshold value){
            put \( \text{sonURL} \) into \( \text{fatherQueue} \); }
        put pages of \( \text{fatherQueue} \) into \( \text{led} \); }
    }
    return \( \text{led} \);
}

**Implement of LED Optical Design Focused Crawler Algorithm**

Led optical design topic crawler algorithm implemented by Java open source web crawler system heritrix. Led optical design progress information is automatically collected, the titles and urls of the information are saved into database. This class is stored in extractorforled.java file.the main method is shown as belows.

```java
public void extract(CrawlURI curi) {
    // get current page reliance score
    float father_score = curi.getUURI().getInheritedScore();
    float child_score = 0;
    // define url reliance score
    o = ParserPage.extractPage(baos.toByteArray(), encoding);
    // get content and url
    if (o != null) {
        urls = (Map<String, String>) o[0];
        // save anchor text and url
        HTMLContent = (String) o[1];
        // calculate current page reliance score
        float sim_father = ComputeCorrelativety.analysis(HTMLContent);
        //handle url according to focused crawler algorithm (Topic-First)
        final Iterator<String> key = urls.keySet().iterator();
        while (key.hasNext()) {
            final String anchor = key.next();
            String link = urls.get(anchor);
            if (anchor != null && link != null) {
                link = getAbsoluteURL(currentUrl, link);
                if (link != null && checkURL(link) == true) {
                    //judge url contains topic dictionary
                    if (JudgeURL.validURL(link) == true) {
                        //...}
                    }
                }
            }
        }
    }
}
```
addLinkFromString(curi, link, "", Link.NAVLINK_HOP, father_score); }
else {
    // calculate url reliance score
    if (sim_father > C) {
        child_score = C * sim_father;
    }
    else {
        child_score = C * father_score;
    }
    // calculate sim(q,anchor) of anchor text and topic q
    float sim_anchor = ComputeCorrelationAnalysis(anchor);
    // calculate reliance score
    if (potential_score >= Threshold) {
        // judge
        addLinkFromString(curi, link, "", Link.Navlink_Hop, child_score); }
    }
}

System Implement and Deployment

Asp.net is used to develop foreground and background system, Java is used to implement led optical design theme crawler system. A large number of controls, such as FileUpload, Repeater, AspNetPager, CKEditor, Heritrix, etc. are used. It also uses web services and ajax. The homepage is shown in Fig.2. Users can visit it to get their needs.

![Figure 2. The home page.](image)

The system is deployed on wuyi university network center server clusters, access address is [http://202.192.240.10:88/](http://202.192.240.10:88/).
Conclusion

We have developed led topic dictionary and the system is fully tested. The function of the system with friendly interface operates normally. Especially the experiments of testing the focused crawler algorithm are carried out under the conditions of same seed number. Its accuracy rate is 13 percentages higher than the breath-first algorithm. The system will be optimized and improved in the process of trial operation in the future.

References