PROYECTO BUSCADOR AED1

Ángel Ruiz Fernandez & Carla Ramos García

12:51 Nov 27, 2024

PROYECTO BUSCADOR AED1: 6 files PAGE

```
 1 #pragma once
 2
 3 #include <string>
 4 #include <vector>
 5 #include <set>
 6 #include <list>
 7 #include <map>
 8 #include <memory>
 9
10 class Pagina {
11     private:
12     std::wstring url, titulo;
13     int relevancia;
14     std::vector<std::wstring> contenido;
15
16     public:
17     Pagina(const std::wstring& _url, const std::wstring& _titulo, int _rel,
18         const std::vector<std::wstring>& _cont);
19     const std::wstring& getUrl() const;
20     const std::wstring& getTitulo() const;
21     void setTitulo(const std::wstring& titulo);
22     int getRelevancia() const;
23     void setRelevancia(int rel);
24     int getContSize() const;
25     const std::vector<std::wstring>& getContenido() const;
26     void setContenido(const std::vector<std::wstring>& palabras);
27 };
28
29 //bool comparar_pagref(const std::list<Pagina>::iterator& l,
30 //    const std::list<Pagina>::iterator& r);
31
32 class PagListIt : public std::list<Pagina>::iterator {
33 public:
34     PagListIt(std::list<Pagina>::iterator& it);
35     bool operator<(const PagListIt& right) const;
36 };
37
38 struct nodo_trie_t {
39     std::set<PagListIt> paginas;
40     std::map<wchar_t, nodo_trie_t> hijos;
41 };
42
43 class Arbol {
44     private:
45     std::map<wchar_t, nodo_trie_t> raiz;
46
47     public:
48     void insertar(const std::wstring& palabra,
49         PagListIt paginaref);
50     std::set<PagListIt> buscar(const std::wstring& palabra);
51     std::vector<std::pair<std::wstring, int>>
52         palabrasPrefijo(const std::wstring& palabra);
53 };
54
55 class Diccionario {
56     private:
57     static constexpr int N = 20000;
58     std::list<Pagina> tabla[N];
59     Arbol arbol;
60     size_t hash(const std::wstring& key);
61     int size = 0;
62
```

```
63      public:
64      void insertar(const Pagina& p);
65      std::vector<Pagina> consultar(const std::wstring& url);
66      std::set<PagListIt> buscarPalabra(const std::wstring& palabra);
67      void pseudoDestructor();
68      size_t getTam();
69      std::vector<std::pair<std::wstring, int>>
70          palabrasPrefijo(const std::wstring& palabra);
71  };
72
```

```
 1 #pragma once
 2
 3 #include <string>
 4 #include <vector>
 5
 6 #include "diccionario.hpp"
 7
 8 auto normalizarc(wchar_t c);
 9 auto normalizarstr(const std::wstring& s);
10 Pagina leerpagina();
11 std::vector<std::wstring> leerpalabras();
12
13 int interpretar(char cmd, Diccionario& dic);
14
```

```cpp
#include "diccionario.hpp"

#include <algorithm>
#include <iostream>
#include <memory>

// Pagina
Pagina::Pagina(const std::wstring& _url, const std::wstring& _titulo, int _rel,
    const std::vector<std::wstring>& _cont)
{
    url = _url;
    titulo = _titulo;
    relevancia = _rel;
    contenido = _cont;
}

const std::wstring& Pagina::getUrl() const {
    return url;
}

const std::wstring& Pagina::getTitulo() const {
    return titulo;
}

void Pagina::setTitulo(const std::wstring& titulo) {
    this->titulo = titulo;
}

int Pagina::getRelevancia() const {
    return relevancia;
}

void Pagina::setRelevancia(int rel) {
    this->relevancia = rel;
}

int Pagina::getContSize() const {
    return contenido.size();
}

const std::vector<std::wstring>& Pagina::getContenido() const {
    return contenido;
}

void Pagina::setContenido(const std::vector<std::wstring>& palabras) {
    contenido = palabras;
}


// Clase referencia a pagina

PagListIt::PagListIt(std::list<Pagina>::iterator& it)
    : std::list<Pagina>::iterator(it)
{

}

bool PagListIt::operator<(const PagListIt& r) const {
    if ((*this)->getRelevancia() != r->getRelevancia())
        return (*this)->getRelevancia() > r->getRelevancia();
    else return (*this)->getUrl() < r->getUrl();
}
```

```
 63
 64 // Diccionario
 65
 66 size_t Diccionario::hash(const std::wstring& key) {
 67     size_t t = 5381;
 68     for (auto c : key)
 69         t = ((t << 5) + t) + c;
 70     return t % N;
 71 }
 72
 73 void insertar_palabras(const std::vector<std::wstring>& palabras,
 74     const std::list<Pagina>::iterator& pagref, Arbol& arbol)
 75 {
 76     for (const std::wstring& p : palabras) {
 77         arbol.insertar(p, (const PagListIt&)pagref);
 78     }
 79 }
 80
 81 void Diccionario::insertar(const Pagina& np) {
 82     auto nhash = hash(np.getUrl());
 83
 84     std::list<Pagina>::iterator it;
 85
 86     auto& vec = tabla[nhash];
 87     for (it = vec.begin(); it != vec.end(); it++) {
 88         if (it->getUrl() == np.getUrl()) {
 89             auto& p = *it;
 90             p.setTitulo(np.getTitulo());
 91             p.setRelevancia(np.getRelevancia());
 92             p.setContenido(np.getContenido());
 93             insertar_palabras(np.getContenido(), it, arbol);
 94             return;
 95         }
 96     }
 97
 98     it = tabla[nhash].insert(tabla[nhash].end(), Pagina(np));
 99     insertar_palabras(np.getContenido(), it, arbol);
100     size++;
101 }
102
103 std::vector<Pagina> Diccionario::consultar(const std::wstring& url) {
104     std::vector<Pagina> resultado;
105     for (const auto& p : tabla[hash(url)])
106         if (p.getUrl() == url)
107             resultado.push_back(p);
108     return resultado;
109 }
110
111 std::set<PagListIt>
112 Diccionario::buscarPalabra(const std::wstring& palabra) {
113     return arbol.buscar(palabra);
114 }
115
116 void Diccionario::pseudoDestructor() {
117     for (int i = 0; i < N; i++)
118         tabla[i].clear();
119 }
120
121 size_t Diccionario::getTam() {
122     return size;
123 }
124
```

```
125 std::vector<std::pair<std::wstring, int>>
126 Diccionario::palabrasPrefijo(const std::wstring& palabra) {
127     return arbol.palabrasPrefijo(palabra);
128 }
129
130 // Arbol
131
132 bool comparar_pagref(const PagListIt& l,
133     const PagListIt& r)
134 {
135     if (l->getRelevancia() != r->getRelevancia())
136         return l->getRelevancia() > r->getRelevancia();
137     else return l->getUrl() < r->getUrl();
138 }
139
140 void Arbol::insertar(const std::wstring& palabra,
141     PagListIt paginaref)
142 {
143     auto subarbol = &raiz; // puntero porque referencia rebindeable
144     std::map<wchar_t, nodo_trie_t>::iterator it;
145
146     for (wchar_t c : palabra) {
147         auto nuevonodo = nodo_trie_t {};
148         // solo inserta si c no existe
149         it = subarbol->insert({c, nuevonodo}).first;
150         subarbol = &it->second.hijos;
151     }
152
153     if (std::find(it->second.paginas.begin(),
154         it->second.paginas.end(), paginaref) == it->second.paginas.end())
155     {
156         it->second.paginas.insert(paginaref);
157     }
158 }
159
160 // no const ref return porque si no se encuentra resultado,
161 // retorna nuevo vector vacio
162 std::set<PagListIt>
163 Arbol::buscar(const std::wstring& palabra) {
164     auto subarbol = &raiz;
165     std::map<wchar_t, nodo_trie_t>::iterator it;
166
167     for (wchar_t c : palabra) {
168         auto nuevonodo = nodo_trie_t();
169         it = subarbol->find(c);
170         if (it == subarbol->end())
171             // retornar vector vacio si no hay resultados
172             return std::set<PagListIt>();
173
174         subarbol = &it->second.hijos;
175     }
176
177     return it->second.paginas;
178 }
179
180 void palabrasPrefijoRecursar(const std::map<wchar_t, nodo_trie_t>& a,
181     std::wstring p, std::vector<std::pair<std::wstring, int>>& palabras)
182 {
183     for (auto n : a) {
184         if (n.second.paginas.size() > 0)
185             palabras.push_back({p + n.first, n.second.paginas.size()});
186         palabrasPrefijoRecursar(n.second.hijos, p + n.first, palabras);
```

```
187     }
188 }
189
190 bool comparadorPalabrasPrefijo(const std::pair<std::wstring, int>& l,
191     const std::pair<std::wstring, int>& r)
192 {
193     if (l.second != r.second) return l.second > r.second;
194     else return l.first < r.first;
195 }
196
197 std::vector<std::pair<std::wstring, int>>
198 Arbol::palabrasPrefijo(const std::wstring& prefijo) {
199     auto subarbol = &raiz;
200     std::map<wchar_t, nodo_trie_t>::iterator it;
201
202     for (wchar_t c : prefijo) {
203         auto nuevonodo = nodo_trie_t();
204         it = subarbol->find(c);
205         if (it == subarbol->end())
206             return std::vector<std::pair<std::wstring, int>>();
207
208         subarbol = &it->second.hijos;
209     }
210
211     std::vector<std::pair<std::wstring, int>> palabras;
212
213     if (it->second.paginas.size() > 0)
214         palabras.push_back({prefijo, it->second.paginas.size()});
215     palabrasPrefijoRecursar(*subarbol, prefijo, palabras);
216     std::sort(palabras.begin(), palabras.end(), comparadorPalabrasPrefijo);
217     return palabras;
218 }
219
```

```cpp
#include "interprete.hpp"

#include <iostream>
#include <limits>
#include <sstream>
#include <algorithm>
#include <iterator>

#include "diccionario.hpp"

auto normalizarc(wchar_t c) {
    c = std::tolower(c); // convierte solo ASCII-7

    if (c == L'Ñ') c = L'ñ';

    if (c == L'Á') c = 'a';
    if (c == L'É') c = 'e';
    if (c == L'Í') c = 'i';
    if (c == L'Ó') c = 'o';
    if (c == L'Ú') c = 'u';
    if (c == L'Ü') c = 'u';

    if (c == L'á') c = 'a';
    if (c == L'é') c = 'e';
    if (c == L'í') c = 'i';
    if (c == L'ó') c = 'o';
    if (c == L'ú') c = 'u';
    if (c == L'ü') c = 'u';
    return c;
}

auto normalizarstr(const std::wstring& s) {
    std::wstring o;
    for (auto c : s)
        o += normalizarc(c);
    return o;
}

Pagina leerpagina() {
    int rel;
    std::wstring url, titulo;
    std::vector<std::wstring> contenido;

    std::wcin >> rel;
    std::wcin >> url;
    std::wcin.ignore(std::numeric_limits<std::streamsize>::max(), '\n');
    std::getline(std::wcin, titulo);

    std::wstring in;
    while ((std::wcin >> in) && (normalizarstr(in) != L"findepagina"))
        contenido.push_back(normalizarstr(in));

    return Pagina(url, titulo, rel, contenido);
}

std::vector<std::wstring> leerpalabras() {
    std::vector<std::wstring> palabras;
    std::wstring palabrasstr;
    std::getline(std::wcin, palabrasstr);

    palabrasstr = normalizarstr(palabrasstr);

```

```
63      auto ss = std::wstringstream(palabrasstr);
64      std::wstring palabra;
65
66      while (ss >> palabra)
67          palabras.push_back(palabra);
68
69      return palabras;
70  }
71
72  int interpretar(char cmd, Diccionario& dic) {
73      switch (cmd) {
74          case 'i': {
75              Pagina np = leerpagina();
76              dic.insertar(np);
77              std::wcout << dic.getTam() << ". " << np.getUrl() << ", "
78                  << np.getTitulo() << ", Rel. " << np.getRelevancia()
79                  << std::endl << np.getContSize() << " palabras" << std::endl;
80          } break;
81          case 'u': {
82              std::wstring url;
83              std::wcin >> url;
84
85              auto p = dic.consultar(url);
86
87              std::wcout << "u " << url << std::endl;
88              for (size_t i = 0; i < p.size(); i++)
89                  std::wcout << i + 1 << ". " << p[i].getUrl() << ", "
90                      << p[i].getTitulo() << ", Rel. " << p[i].getRelevancia()
91                      << std::endl;
92              std::wcout << "Total: " << p.size() << " resultados" << std::endl;
93
94          } break;
95          case 'b': {
96              std::wstring palabra;
97              std::wcin >> palabra;
98              palabra = normalizarstr(palabra);
99
100             auto paginas = dic.buscarPalabra(palabra);
101
102             std::wcout << "b " << palabra << std::endl;
103             int i = 1;
104             for (auto& p : paginas) {
105                 std::wcout << i << ". " <<  p->getUrl() << ", "
106                     << p->getTitulo() << ", Rel. "
107                     << p->getRelevancia() << std::endl;
108                 i++;
109             }
110             std::wcout << "Total: " << paginas.size() << " resultados"
111                 << std::endl;
112         } break;
113         case 'a': {
114             std::vector<std::wstring> palabras = leerpalabras();
115
116             std::wcout << "a";
117             for (auto p : palabras)
118                 std::wcout << " " << p;
119             std::wcout << std::endl;
120
121             std::set<PagListIt> interseccion;
122
123             if (!palabras.empty()) {
124                 interseccion = dic.buscarPalabra(palabras[0]);
```

```
125                 palabras.erase(palabras.begin());
126             }
127 
128             for (auto& p : palabras) {
129                 std::set<PagListIt> paginas = dic.buscarPalabra(p), nuevo;
130 
131                 std::set_intersection(interseccion.begin(), interseccion.end(),
132                     paginas.begin(), paginas.end(),
133                     std::inserter(nuevo, nuevo.begin()));
134 
135                 interseccion = nuevo;
136             }
137 
138             int i = 1;
139             for (auto& p : interseccion) {
140                 std::wcout << i << ". " <<  p->getUrl() << ", "
141                     << p->getTitulo() << ", Rel. "
142                     << p->getRelevancia() << std::endl;
143                 i++;
144             }
145             std::wcout << "Total: " << interseccion.size()
146                 << " resultados" << std::endl;
147         } break;
148         case 'o': {
149             std::vector<std::wstring> palabras = leerpalabras();
150 
151             std::wcout << "o";
152             for (auto p : palabras)
153                 std::wcout << " " << p;
154             std::wcout << std::endl;
155 
156             std::set<PagListIt> unionp;
157 
158             for (auto& p : palabras) {
159                 std::set<PagListIt> paginas = dic.buscarPalabra(p), nuevo;
160 
161                 std::set_union(unionp.begin(), unionp.end(),
162                     paginas.begin(), paginas.end(),
163                     std::inserter(unionp, unionp.begin()));
164             }
165 
166             int i = 1;
167             for (auto& p : unionp) {
168                 std::wcout << i << ". " <<  p->getUrl() << ", "
169                     << p->getTitulo() << ", Rel. "
170                     << p->getRelevancia() << std::endl;
171                 i++;
172             }
173 
174             std::wcout << "Total: " << unionp.size() << " resultados" << std::endl;
175         } break;
176         case 'p': {
177             std::wstring prefijo;
178             std::wcin >> prefijo;
179             prefijo = normalizarstr(prefijo);
180 
181             auto palabras = dic.palabrasPrefijo(prefijo);
182 
183             std::wcout << "p " << prefijo << std::endl;
184 
185             int i = 1;
186             for (auto& p : palabras) {
```

```
187                 std::wcout << i << ". " << p.first << ", " << p.second
188                     << std::endl;
189                 i++;
190             }
191 
192             std::wcout << "Total: " << palabras.size()
193                 << " resultados" << std::endl;
194         } break;
195         case 's': {
196             std::wcout << "Saliendo..." << std::endl;
197             return 0;
198         } break;
199         default: return -1;
200     }
201     return 0;
202 }
203 
```

```cpp
#include <iostream>
#include <vector>
#include <string>
#include <sstream>
#include <limits>

#include "diccionario.hpp"
#include "interprete.hpp"

int main() {
    // trabajar con UTF-8 y caracteres anchos
    std::setlocale(LC_ALL, "es_ES.UTF-8");

    Diccionario dic;

    std::wstring in;
    while (std::wcin >> in) {
        if (in.length() != 1) continue;
        int r = interpretar(in[0], dic);
        if (r < 0) return 1;
    }
}

```

```
PROJECT := buscador
BINARY := a.out
CXX := g++
CXXFLAGS := --std=c++17 -Wall -pedantic -g -O0
LDFLAGS :=

SRC := $(wildcard *.cpp)
OBJ := $(patsubst %.cpp,%.o,$(SRC))

all: $(BINARY)

$(BINARY): $(OBJ)
    $(CXX) -o $(BINARY) $(OBJ) $(LDFLAGS)

%.o: %.cpp %.hpp
    $(CXX) -c $(CXXFLAGS) $<

.PHONY: test
test: $(BINARY)
    ./$(BINARY) < test_stdin.txt > run_stdout.txt
    cmp -s test_stdout.txt run_stdout.txt && echo "PASS" || echo "FAIL"

.PHONY: clean
clean:
    rm $(BINARY) *.o run* *.tar

.PHONY: tar
tar: $(PROJECT).tar

$(PROJECT).tar: $(SRC) Makefile
    tar cf $(PROJECT).tar *.cpp *.hpp Makefile

.PHONY: submit
submit: tar
    python3 submit.py
```