

Порівняльний аналіз методів обробки тексту: C (string.h), C++ (string) та C++11 (regex)

Анотація

Цей реферат проводить порівняльний аналіз трьох різних підходів до вирішення чотирьох практичних задач обробки тексту: пошук гіперпосилань, конвертація форматів дат, форматування чисел та заміна на поточну дату/час. Реалізації розглядаються у трьох технологічних контекстах: чистий С з використанням бібліотеки `string.h` (файл `a_tasks.cpp`), C++ з бібліотекою `<string>` та ручним парсингом (файл `b_string.cpp`), та сучасний C++11 з використанням бібліотеки регулярних виразів `<regex>` (файл `c_regex.cpp`). Аналіз фокусується на порівнянні читабельності коду, ефективності, гнучкості та придатності кожного підходу для вирішення типових задач текстових трансформацій.

Зміст

1 Вступ	2
2 Огляд реалізацій	2
2.1 Файл <code>a_tasks.cpp</code> : Процедурний стиль на С	2
2.1.1 Ключові функції бібліотеки <code>string.h</code> та власні парсери	2
2.2 Файл <code>b_string.cpp</code> : Об'єктно-орієнтований підхід на C++ з ручним парсингом	3
2.2.1 Ключові методи та логіка	3
2.3 Файл <code>c_regex.cpp</code> : Декларативний підхід на C++11 з регулярними виразами	4
2.3.1 Ключові регулярні вирази та класи	4
3 Порівняльний аналіз підходів	5
3.1 Порівняльна таблиця характеристик підходів	5
3.2 Детальний аналіз окремих аспектів	5
3.2.1 Обробка помилок та надійність	5
3.2.2 Продуктивність у контексті задач	5
4 Висновки та рекомендації	6
4.1 Сфери застосування	6
4.2 Загальні висновки	7

1 Вступ

Обробка тексту є фундаментальною задачею в програмуванні, що зустрічається в різних доменах: від систем адміністрування до веб-скрапінгу та аналізу даних. У даному проекті реалізовано набір з чотирьох типових задач:

1. **Задача 1:** Виявлення гіперпосилань (URL) у тексті та визначення слова, що безпосередньо передує кожному посиланню.
2. **Задача 2:** Конвертація дат із формату ДД/ММ/РР або ДД/ММ/РРРР у текстово-числовий формат (РРРР) НазваМісяця (ДД) українською мовою.
3. **Задача 3:** Пошук усіх чисел (цілих, дійсних, наукового запису) у тексті та їх форматування у вигляді (значення.з.4.знаками.після.крапки).
4. **Задача 4:** Заміна всіх знайдених у тексті дат у форматі результату Задачі 2 та часів у форматі ГГ:ХХ:СС на поточну дату та час системи.

Для вирішення цих задач розроблено три альтернативні реалізації, що демонструють еволюцію інструментів обробки рядків:

- **Підхід А (процедурний, низькорівневий):** Реалізація на чистій мові С з використанням бібліотек `string.h`, `ctype.h` та власних функцій парсингу (файл `a_tasks.cpp`).
- **Підхід В (об'єктно-орієнтований, ручний парсинг):** Реалізація на С++ з використанням класу `std::string` та бібліотеки `<algorithm>`, але без регулярних виразів (файл `b_string.cpp`).
- **Підхід С (сучасний, декларативний):** Реалізація на С++11 з використанням бібліотеки регулярних виразів `<regex>` для компактного та виразного опису патернів пошуку (файл `c_regex.cpp`).

Метою роботи є не лише демонстрація рішень, але й системний порівняльний аналіз їхніх архітектурних особливостей, продуктивності та сфер застосування.

2 Огляд реалізацій

2.1 Файл `a_tasks.cpp`: Процедурний стиль на С

Ця реалізація характеризується низькорівневим доступом до пам'яті, явним управлінням ресурсами та використанням функцій стандартної бібліотеки С для маніпуляцій з рядками.

2.1.1 Ключові функції бібліотеки `string.h` та власні парсери

- `strstr()`, `strlen()`, `memcpuy()` – базові операції з рядками.
- Власна функція `parse_date_at()` для розпізнавання дат у різних форматах через послідовне читання символів.

- Власна функція `parse_number_at()` для вилучення чисел з тексту, що враховує знаки, десяткові роздільники (крапку/кому) та експоненціальну нотацію.
- Функції `task1_links()`, `task2_find_and_print_dates()`, `task3_find_and_print_numbers()` та `task4_replace_dates_and_write()` інкапсулюють логіку кожної задачі, але працюють на рівні сиріх вказівників та довжин.

Характерний фрагмент коду (парсинг дати):

```

1 static int parse_date_at(const char* s, size_t len, size_t pos,
2                           size_t* out_start, size_t* out_end,
3                           int* out_d, int* out_m, int* out_y) {
4     // ...
5     //           i           i           i   ,
6     //           i           i           i   ,
7     //           i
8     //           i
9     //           i
10    //           i
11    //           i
12    //           i
13    //           i
14    //           i
15    //           i
16    //           i
17    //           i
18    //           i
19    //           i
20    //           i
21    //           i
22    //           i
23    //           i
24    //           i
25    //           i
26    //           i
27    //           i
28    //           i
29    //           i
30    //           i
31    //           i
32    //           i
33    //           i
34    //           i
35    //           i
36    //           i
37    //           i
38    //           i
39    //           i
40    //           i
41    //           i
42    //           i
43    //           i
44    //           i
45    //           i
46    //           i
47    //           i
48    //           i
49    //           i
50    //           i
51    //           i
52    //           i
53    //           i
54    //           i
55    //           i
56    //           i
57    //           i
58    //           i
59    //           i
60    //           i
61    //           i
62    //           i
63    //           i
64    //           i
65    //           i
66    //           i
67    //           i
68    //           i
69    //           i
70    //           i
71    //           i
72    //           i
73    //           i
74    //           i
75    //           i
76    //           i
77    //           i
78    //           i
79    //           i
80    //           i
81    //           i
82    //           i
83    //           i
84    //           i
85    //           i
86    //           i
87    //           i
88    //           i
89    //           i
90    //           i
91    //           i
92    //           i
93    //           i
94    //           i
95    //           i
96    //           i
97    //           i
98    //           i
99    //           i
100   //           i
101   //           i
102   //           i
103   //           i
104   //           i
105   //           i
106   //           i
107   //           i
108   //           i
109   //           i
110   //           i
111   //           i
112   //           i
113   //           i
114   //           i
115   //           i
116   //           i
117   //           i
118   //           i
119   //           i
120   //           i
121   //           i
122   //           i
123   //           i
124   //           i
125   //           i
126   //           i
127   //           i
128   //           i
129   //           i
130   //           i
131   //           i
132   //           i
133   //           i
134   //           i
135   //           i
136   //           i
137   //           i
138   //           i
139   //           i
140   //           i
141   //           i
142   //           i
143   //           i
144   //           i
145   //           i
146   //           i
147   //           i
148   //           i
149   //           i
150   //           i
151   //           i
152   //           i
153   //           i
154   //           i
155   //           i
156   //           i
157   //           i
158   //           i
159   //           i
160   //           i
161   //           i
162   //           i
163   //           i
164   //           i
165   //           i
166   //           i
167   //           i
168   //           i
169   //           i
170   //           i
171   //           i
172   //           i
173   //           i
174   //           i
175   //           i
176   //           i
177   //           i
178   //           i
179   //           i
180   //           i
181   //           i
182   //           i
183   //           i
184   //           i
185   //           i
186   //           i
187   //           i
188   //           i
189   //           i
190   //           i
191   //           i
192   //           i
193   //           i
194   //           i
195   //           i
196   //           i
197   //           i
198   //           i
199   //           i
200   //           i
201   //           i
202   //           i
203   //           i
204   //           i
205   //           i
206   //           i
207   //           i
208   //           i
209   //           i
210   //           i
211   //           i
212   //           i
213   //           i
214   //           i
215   //           i
216   //           i
217   //           i
218   //           i
219   //           i
220   //           i
221   //           i
222   //           i
223   //           i
224   //           i
225   //           i
226   //           i
227   //           i
228   //           i
229   //           i
230   //           i
231   //           i
232   //           i
233   //           i
234   //           i
235   //           i
236   //           i
237   //           i
238   //           i
239   //           i
240   //           i
241   //           i
242   //           i
243   //           i
244   //           i
245   //           i
246   //           i
247   //           i
248   //           i
249   //           i
250   //           i
251   //           i
252   //           i
253   //           i
254   //           i
255   //           i
256   //           i
257   //           i
258   //           i
259   //           i
260   //           i
261   //           i
262   //           i
263   //           i
264   //           i
265   //           i
266   //           i
267   //           i
268   //           i
269   //           i
270   //           i
271   //           i
272   //           i
273   //           i
274   //           i
275   //           i
276   //           i
277   //           i
278   //           i
279   //           i
280   //           i
281   //           i
282   //           i
283   //           i
284   //           i
285   //           i
286   //           i
287   //           i
288   //           i
289   //           i
290   //           i
291   //           i
292   //           i
293   //           i
294   //           i
295   //           i
296   //           i
297   //           i
298   //           i
299   //           i
300   //           i
301   //           i
302   //           i
303   //           i
304   //           i
305   //           i
306   //           i
307   //           i
308   //           i
309   //           i
310   //           i
311   //           i
312   //           i
313   //           i
314   //           i
315   //           i
316   //           i
317   //           i
318   //           i
319   //           i
320   //           i
321   //           i
322   //           i
323   //           i
324   //           i
325   //           i
326   //           i
327   //           i
328   //           i
329   //           i
330   //           i
331   //           i
332   //           i
333   //           i
334   //           i
335   //           i
336   //           i
337   //           i
338   //           i
339   //           i
340   //           i
341   //           i
342   //           i
343   //           i
344   //           i
345   //           i
346   //           i
347   //           i
348   //           i
349   //           i
350   //           i
351   //           i
352   //           i
353   //           i
354   //           i
355   //           i
356   //           i
357   //           i
358   //           i
359   //           i
360   //           i
361   //           i
362   //           i
363   //           i
364   //           i
365   //           i
366   //           i
367   //           i
368   //           i
369   //           i
370   //           i
371   //           i
372   //           i
373   //           i
374   //           i
375   //           i
376   //           i
377   //           i
378   //           i
379   //           i
380   //           i
381   //           i
382   //           i
383   //           i
384   //           i
385   //           i
386   //           i
387   //           i
388   //           i
389   //           i
390   //           i
391   //           i
392   //           i
393   //           i
394   //           i
395   //           i
396   //           i
397   //           i
398   //           i
399   //           i
400   //           i
401   //           i
402   //           i
403   //           i
404   //           i
405   //           i
406   //           i
407   //           i
408   //           i
409   //           i
410   //           i
411   //           i
412   //           i
413   //           i
414   //           i
415   //           i
416   //           i
417   //           i
418   //           i
419   //           i
420   //           i
421   //           i
422   //           i
423   //           i
424   //           i
425   //           i
426   //           i
427   //           i
428   //           i
429   //           i
430   //           i
431   //           i
432   //           i
433   //           i
434   //           i
435   //           i
436   //           i
437   //           i
438   //           i
439   //           i
440   //           i
441   //           i
442   //           i
443   //           i
444   //           i
445   //           i
446   //           i
447   //           i
448   //           i
449   //           i
450   //           i
451   //           i
452   //           i
453   //           i
454   //           i
455   //           i
456   //           i
457   //           i
458   //           i
459   //           i
460   //           i
461   //           i
462   //           i
463   //           i
464   //           i
465   //           i
466   //           i
467   //           i
468   //           i
469   //           i
470   //           i
471   //           i
472   //           i
473   //           i
474   //           i
475   //           i
476   //           i
477   //           i
478   //           i
479   //           i
480   //           i
481   //           i
482   //           i
483   //           i
484   //           i
485   //           i
486   //           i
487   //           i
488   //           i
489   //           i
490   //           i
491   //           i
492   //           i
493   //           i
494   //           i
495   //           i
496   //           i
497   //           i
498   //           i
499   //           i
500   //           i
501   //           i
502   //           i
503   //           i
504   //           i
505   //           i
506   //           i
507   //           i
508   //           i
509   //           i
510   //           i
511   //           i
512   //           i
513   //           i
514   //           i
515   //           i
516   //           i
517   //           i
518   //           i
519   //           i
520   //           i
521   //           i
522   //           i
523   //           i
524   //           i
525   //           i
526   //           i
527   //           i
528   //           i
529   //           i
530   //           i
531   //           i
532   //           i
533   //           i
534   //           i
535   //           i
536   //           i
537   //           i
538   //           i
539   //           i
540   //           i
541   //           i
542   //           i
543   //           i
544   //           i
545   //           i
546   //           i
547   //           i
548   //           i
549   //           i
550   //           i
551   //           i
552   //           i
553   //           i
554   //           i
555   //           i
556   //           i
557   //           i
558   //           i
559   //           i
560   //           i
561   //           i
562   //           i
563   //           i
564   //           i
565   //           i
566   //           i
567   //           i
568   //           i
569   //           i
570   //           i
571   //           i
572   //           i
573   //           i
574   //           i
575   //           i
576   //           i
577   //           i
578   //           i
579   //           i
580   //           i
581   //           i
582   //           i
583   //           i
584   //           i
585   //           i
586   //           i
587   //           i
588   //           i
589   //           i
590   //           i
591   //           i
592   //           i
593   //           i
594   //           i
595   //           i
596   //           i
597   //           i
598   //           i
599   //           i
600   //           i
601   //           i
602   //           i
603   //           i
604   //           i
605   //           i
606   //           i
607   //           i
608   //           i
609   //           i
610   //           i
611   //           i
612   //           i
613   //           i
614   //           i
615   //           i
616   //           i
617   //           i
618   //           i
619   //           i
620   //           i
621   //           i
622   //           i
623   //           i
624   //           i
625   //           i
626   //           i
627   //           i
628   //           i
629   //           i
630   //           i
631   //           i
632   //           i
633   //           i
634   //           i
635   //           i
636   //           i
637   //           i
638   //           i
639   //           i
640   //           i
641   //           i
642   //           i
643   //           i
644   //           i
645   //           i
646   //           i
647   //           i
648   //           i
649   //           i
650   //           i
651   //           i
652   //           i
653   //           i
654   //           i
655   //           i
656   //           i
657   //           i
658   //           i
659   //           i
660   //           i
661   //           i
662   //           i
663   //           i
664   //           i
665   //           i
666   //           i
667   //           i
668   //           i
669   //           i
670   //           i
671   //           i
672   //           i
673   //           i
674   //           i
675   //           i
676   //           i
677   //           i
678   //           i
679   //           i
680   //           i
681   //           i
682   //           i
683   //           i
684   //           i
685   //           i
686   //           i
687   //           i
688   //           i
689   //           i
690   //           i
691   //           i
692   //           i
693   //           i
694   //           i
695   //           i
696   //           i
697   //           i
698   //           i
699   //           i
700   //           i
701   //           i
702   //           i
703   //           i
704   //           i
705   //           i
706   //           i
707   //           i
708   //           i
709   //           i
710   //           i
711   //           i
712   //           i
713   //           i
714   //           i
715   //           i
716   //           i
717   //           i
718   //           i
719   //           i
720   //           i
721   //           i
722   //           i
723   //           i
724   //           i
725   //           i
726   //           i
727   //           i
728   //           i
729   //           i
730   //           i
731   //           i
732   //           i
733   //           i
734   //           i
735   //           i
736   //           i
737   //           i
738   //           i
739   //           i
740   //           i
741   //           i
742   //           i
743   //           i
744   //           i
745   //           i
746   //           i
747   //           i
748   //           i
749   //           i
750   //           i
751   //           i
752   //           i
753   //           i
754   //           i
755   //           i
756   //           i
757   //           i
758   //           i
759   //           i
760   //           i
761   //           i
762   //           i
763   //           i
764   //           i
765   //           i
766   //           i
767   //           i
768   //           i
769   //           i
770   //           i
771   //           i
772   //           i
773   //           i
774   //           i
775   //           i
776   //           i
777   //           i
778   //           i
779   //           i
780   //           i
781   //           i
782   //           i
783   //           i
784   //           i
785   //           i
786   //           i
787   //           i
788   //           i
789   //           i
790   //           i
791   //           i
792   //           i
793   //           i
794   //           i
795   //           i
796   //           i
797   //           i
798   //           i
799   //           i
800   //           i
801   //           i
802   //           i
803   //           i
804   //           i
805   //           i
806   //           i
807   //           i
808   //           i
809   //           i
810   //           i
811   //           i
812   //           i
813   //           i
814   //           i
815   //           i
816   //           i
817   //           i
818   //           i
819   //           i
820   //           i
821   //           i
822   //           i
823   //           i
824   //           i
825   //           i
826   //           i
827   //           i
828   //           i
829   //           i
830   //           i
831   //           i
832   //           i
833   //           i
834   //           i
835   //           i
836   //           i
837   //           i
838   //           i
839   //           i
840   //           i
841   //           i
842   //           i
843   //           i
844   //           i
845   //           i
846   //           i
847   //           i
848   //           i
849   //           i
850   //           i
851   //           i
852   //           i
853   //           i
854   //           i
855   //           i
856   //           i
857   //           i
858   //           i
859   //           i
860   //           i
861   //           i
862   //           i
863   //           i
864   //           i
865   //           i
866   //           i
867   //           i
868   //           i
869   //           i
870   //           i
871   //           i
872   //           i
873   //           i
874   //           i
875   //           i
876   //           i
877   //           i
878   //           i
879   //           i
880   //           i
881   //           i
882   //           i
883   //           i
884   //           i
885   //           i
886   //           i
887   //           i
888   //           i
889   //           i
890   //           i
891   //           i
892   //           i
893   //           i
894   //           i
895   //           i
896   //           i
897   //           i
898   //           i
899   //           i
900   //           i
901   //           i
902   //           i
903   //           i
904   //           i
905   //           i
906   //           i
907   //           i
908   //           i
909   //           i
910   //           i
911   //           i
912   //           i
913   //           i
914   //           i
915   //           i
916   //           i
917   //           i
918   //           i
919   //           i
920   //           i
921   //           i
922   //           i
923   //           i
924   //           i
925   //           i
926   //           i
927   //           i
928   //           i
929   //           i
930   //           i
931   //           i
932   //           i
933   //           i
934   //           i
935   //           i
936   //           i
937   //           i
938   //           i
939   //           i
940   //           i
941   //           i
942   //           i
943   //           i
944   //           i
945   //           i
946   //           i
947   //           i
948   //           i
949   //           i
950   //           i
951   //           i
952   //           i
953   //           i
954   //           i
955   //           i
956   //           i
957   //           i
958   //           i
959   //           i
960   //           i
961   //           i
962   //           i
963   //           i
964   //           i
965   //           i
966   //           i
967   //           i
968   //           i
969   //           i
970   //           i
971   //           i
972   //           i
973   //           i
974   //           i
975   //           i
976   //           i
977   //           i
978   //           i
979   //           i
980   //           i
981   //           i
982   //           i
983   //           i
984   //           i
985   //           i
986   //           i
987   //           i
988   //           i
989   //           i
990   //           i
991   //           i
992   //           i
993   //           i
994   //           i
995   //           i
996   //           i
997   //           i
998   //           i
999   //           i
1000  //           i
1001  //           i
1002  //           i
1003  //           i
1004  //           i
1005  //           i
1006  //           i
1007  //           i
1008  //           i
1009  //           i
1010  //           i
1011  //           i
1012  //           i
1013  //           i
1014  //           i
1015  //           i
1016  //           i
1017  //           i
1018  //           i
1019  //           i
1020  //           i
1021  //           i
1022  //           i
1023  //           i
1024  //           i
1025  //           i
1026  //           i
1027  //           i
1028  //           i
1029  //           i
1030  //           i
1031  //           i
1032  //           i
1033  //           i
1034  //           i
1035  //           i
1036  //           i
1037  //           i
1038  //           i
1039  //           i
1040  //           i
1041  //           i
1042  //           i
1043  //           i
1044  //           i
1045  //           i
1046  //           i
1047  //           i
1048  //           i
1049  //           i
1050  //           i
1051  //           i
1052  //           i
1053  //           i
1054  //           i
1055  //           i
1056  //           i
1057  //           i
1058  //           i
1059  //           i
1060  //           i
1061  //           i
1062  //           i
1063  //           i
1064  //           i
1065  //           i
1066  //           i
1067  //           i
1068  //           i
1069  //           i
1070  //           i
1071  //           i
1072  //           i
1073  //           i
1074  //           i
1075  //           i
1076  //           i
1077  //           i
1078  //           i
1079  //           i
1080  //           i
1081  //           i
1082  //           i
1083  //           i
1084  //           i
1085  //           i
1086  //           i
1087  //           i
1088  //           i
1089  //           i
1090  //           i
1091  //           i
1092  //           i
1093  //           i
1094  //           i
1095  //           i
1096  //           i
1097  //           i
1098  //           i
1099  //           i
1100  //           i
1101  //           i
1102  //           i
1103  //           i
1104  //           i
1105  //           i
1106  //           i
1107  //           i
1108  //           i
1109  //           i
1110  //           i
1111  //           i
1112  //           i
1113  //           i
1114  //           i
1115  //           i
1116  //           i
1117  //           i
1118  //           i
1119  //           i
1120  //           i
1121  //           i
1122  //           i
1123  //           i
1124  //           i
1125  //           i
1126  //           i
1127  //           i
1128  //           i
1129  //           i
1130  //           i
1131  //           i
1132  //           i
1133  //           i
1134  //           i
1135  //           i
1136  //           i
1137  //           i
1138  //           i
1139  //           i
1140  //           i
1141  //           i
1142  //           i
1143  //           i
1144  //           i
1145  //           i
1146  //           i
1147  //           i
1148  //           i
1149  //           i
1150  //           i
1151  //           i
1152  //           i
1153  //           i
1154  //           i
1155  //           i
1156  //           i
1157  //           i
1158  //           i
1159  //           i
1160  //           i
1161  //           i
1162  //           i
1163  //           i
1164  //           i
1165  //           i
1166  //           i
1167  //           i
1168  //           i
1169  //           i
1170  //           i
1171  //           i
1172  //           i
1173  //           i
1174  //           i
1175  //           i
1176  //           i
1177  //           i
1178  //           i
1179  //           i
1180  //           i
1181  //           i
1182  //           i
1183  //           i
1184  //           i
1185  //           i
1186  //           i
1187  //           i
1188  //           i
1189  //           i
1190  //           i
1191  //           i
1192  //           i
1193  //           i
1194  //           i
1195  //           i
1196  //           i
1197  //           i
1198  //           i
1199  //           i
1200  //           i
1201  //           i
1202  //           i
1203  //           i
1204  //           i
1205  //           i
1206  //           i
1207  //           i
1208  //           i
1209  //           i
1210  //           i
1211  //           i
1212  //           i
1213  //           i
1214  //           i
1215  //           i
1216  //           i
1217  //           i
1218  //           i
1219  //           i
1220  //           i
1221  //           i
1222  //           i
1223  //           i
1224  //           i
1225  //           i
1226  //           i
1227  //           i
1228  //           i
1229  //           i
1230  //           i
1231  //           i
1232  //           i
1233  //           i
1234  //           i
1235  //           i
1236  //           i
1237  //           i
1238  //           i
1239  //           i
1240  //           i
1241  //           i
1242  //           i
1243  //           i
1244  //           i
1245  //           i
1246  //           i
1247  //           i
1248  //           i
1249  //           i
1250  //           i
1251  //           i
1252  //           i
1253  //           i
1254  //           i
1255  //           i
1256  //           i
1257  //           i
1258  //           i
1259  //           i
1260  //           i
1261  //           i
1262  //           i
1263  //           i
1264  //           i
1265  //           i
1266  //           i
1267  //           i
1268  //           i
1269  //           i
1270  //           i
1271  //           i
1272  //           i
1273  //           i
1274  //           i
1275  //           i
1276  //           i
1277  //           i
1278  //           i
1279  //           i
1280  //           i
1281  //           i
1282  //           i
1283  //           i
1284  //           i
1285  //           i
1286  //           i
1287  //           i
1288  //           i
1289  //           i
1290  //           i
1291  //           i
1292  //           i
1293  //           i
1294  //           i
1295  //           i
1296  //           i
1297  //           i
1298  //           i
1299  //           i
1300  //           i
1301  //           i
1302  //           i
1303  //           i
1304  //           i
1305  //           i
1306  //           i
1307  //           i
1308  //           i
1309  //           i
1310  //           i
1311  //           i
1312  //           i
1313  //           i
1314  //           i
1315  //           i
1316  //           i
1317  //           i
1318  //           i
1319  //           i
1320  //           i
1321  //           i
1322  //           i
1323  //           i
1324  //           i
1325  //           i
1326  //           i
1327  //          
```

2.3 Файл с _ regex.cpp: Декларативний підхід на C++11 з регулярними виразами

Ця реалізація максимально використовує сучасні інструменти C++11. Ключовою є бібліотека `<regex>`, що дозволяє описувати шаблони пошуку у декларативному стилі.

2.3.1 Ключові регулярні вирази та класи

- `regex_url_regex(R"((https?://[< >]+))")` – –URL.
- `regex_date_regex(R"((\d{1,2})/(\d{1,2})/(\d{2}(?:\d{2})?))")` – для виявлення дат.
- `regex_number_regex(R"(([+-]?\d+(?:[.,]\d+)?(?:[eE][+-]?\d+)?))")` – для пошуку чисел.
- Використання `sregex_iterator` для ітерації по всіх збігах у тексті.
- Використання `regex_replace()` для глобальної заміни за шаблоном (Задача 4).

Характерний фрагмент коду (конвертація дат за допомогою `regex`):

```
1 regex date_regex(R"((\d{1,2})/(\d{1,2})/(\d{2}(?:\d{2})?))");
2 sregex_iterator date_begin(result.begin(), result.end(),
3   date_regex);
4 for (auto it = date_begin; it != sregex_iterator(); ++it) {
5   smatch match = *it;
6   string day_str = match[1].str();
7   string month_str = match[2].str();
8   string year_str = match[3].str();
9 }
```

3 Порівняльний аналіз підходів

3.1 Порівняльна таблиця характеристик підходів

Критерій	C (a_tasks.cpp)	C++ без regex (b_string.cpp)	C++11 з regex (c_regex.cpp)
Читабельність	Низька - великий обсяг рутинного коду для парсингу	Середня - логіка виражена чіткіше, але парсинг залишається імперативним	Висока - шаблони описані декларативно, логіка перетворень чітка
Гнучкість змін	Дуже низька - зміна формату вимагає переробки парсера	Низька/Середня - зміни потребують переробки алгоритмів	Висока - часто досить змінити регулярний вираз
Продуктивність	Найвища - мінімум накладних витрат, пряма робота з пам'яттю	Середня - виклики методів std::string	Нижча - накладні витрати на компіляцію та виконання regex
Налагодження	Складне - важко відстежити помилки парсингу	Середнє - полегшене структурою C++	Середнє/Складне - налагодження regex може бути складним
Безпека	Низька - ризик перевнення буфера, помилок з вказівниками	Висока - std::string забезпечує безпечне керування пам'яттю	Висока - використання безпечних контейнерів C++
Підтримка Unicode	Фактично відсутня - працює з байтами	Часткова - залежить від налаштувань	Найкраща - можливість використання класів символів
Обсяг коду	Найбільший (понад 300 рядків)	Менший (понад 250 рядків)	Найменший (блізько 200 рядків)

Табл. 1: Порівняльна характеристика трьох підходів до обробки тексту

3.2 Детальний аналіз окремих аспектів

3.2.1 Обробка помилок та надійність

- C підхід:** Потрібна ретельна ручна перевірка всіх граничних умов, довжин буферів та значень, що повертаються (наприклад, NULL від malloc()).
- C++ підхід (без regex):** Безпечніша завдяки RAII (std::string, потоки файлів). Потенційні помилки в логіці ручного парсингу.
- C++ regex підхід:** Бібліотека <regex> викидає винятки при помилках компіляції виразів, що спрощує обробку помилок на етапі ініціалізації. Логічні помилки в шаблонах можуть привести до неочікуваних результатів пошуку.

3.2.2 Продуктивність у контексті задач

- Простий пошук підрядка (URL):** C-підхід з strstr() може бути швидшим, але regex ("https://") більш виразний і коректно обробляє варіативність.

- **Складний структурований парсинг (дати, числа):** Рукописні парсери (C та C++) можуть бути оптимізовані під конкретний формат і виявляються швидшими за regex для фіксованих, відомих заздалегідь шаблонів. Однак regex значно перевершує за швидкістю розробки та гнучкістю.
- **Глобальна заміна (Задача 4):** `regex_replace()` є інструментом найвищого рівня, що робить операцію дуже компактною, але може поступатися за швидкістю спеціалізованому імперативному алгоритму.

4 Висновки та рекомендації

Проведений аналіз дозволяє сформулювати чіткі рекомендації щодо вибору підходу:

4.1 Сфери застосування

1. **Використовувати чистий C (a_tasks.cpp) доцільно, якщо:**
 - Програма працює в середовищах з **жорсткими обмеженнями на ресурси** (пам'ять, процесорний час).
 - Потрібна **максимальна швидкість** виконання на відомих, незмінних форматах даних.
 - Цільова платформа має **обмежену підтримку бібліотек C++** або використовується legacy-кодова база.
 - Розробники мають глибокий досвід і готові до рутинної роботи з пам'яттю та вказівниками.
2. **Використовувати C++ з ручним парсингом (b_string.cpp) рекомендується, якщо:**
 - Задача потребує **складних, нестандартних алгоритмів** обробки, які важко виразити регулярним виразом.
 - **Формати даних стабільні** і не очікуються їхні часті зміни.
 - Потрібний **баланс між продуктивністю та сучасністю кодової бази**.
 - Компілятор не підтримує C++11 повною мірою або використання `<regex>` є небажаним через розмір бібліотеки.
3. **Використовувати C++11 з регулярними виразами (c_regex.cpp) є оптимальним вибором, коли:**
 - Пріоритетом є **швидкість розробки, читабельність та зручність супроводу коду**.
 - **Формати даних можуть змінюватися** або потрібна гнучкість у описі шаблонів пошуку.
 - Задачі пов'язані з **валідацією, виділенням структурованих даних** з тексту (наприклад, логування, парсинг логів, обробка напівструктурзованих даних).
 - Проект використовує сучасний стек C++ (11/14/17 стандарт), і немає критичних обмежень на час виконання, пов'язаних із компіляцією regex.

4.2 Загальні висновки

1. **Еволюція інструментів** від процедурного парсингу на С до декларативних регулярних виразів у C++11 чітко відображає загальну тенденцію в програмуванні: підвищення рівня абстракції для зменшення обсягу коду та покращення його якості.

2. **Немає універсального найкращого рішення.** Кожен підхід має свою «нішу». С-підхід незамінний для системного програмування, С++ з regex — для швидкої розробки бізнес-логіки, пов’язаної з текстом.

3. **Регулярні вирази — це потужний інструмент, але не «срібна куля».** Вони можуть ставати складними для читання та налагодження, а їхня продуктивність не завжди є оптимальною для найпростіших операцій. Критично важливо добре документувати складні regex патерни.

4. **Гіbridний підхід** часто виявляється найпрактичнішим у реальних проектах: використання regex для первинного виділення даних (наприклад, знаходження фрагмента, що містить дату) з подальшим точним парсингом виділеного фрагмента спеціалізованими функціями або бібліотеками.

У цілому, для нових проектів обробки тексту, де немає екстремальних вимог до продуктивності, **підхід з використанням регулярних виразів (C) є рекомендованим.** Він забезпечує найкращу комбінацію розумної швидкості розробки, простоти підтримки та достатньої гнучкості для адаптації до змін у вхідних даних.