

3 Phòng thí nghiệm số 3. "Các biểu thức chính quy và ngôn ngữ đánh dấu tài liệu. Trần"

Để xác định tùy chọn, hãy sử dụng mã số nhân sự của bạn, mã số này có thể tìm thấy trong MIS. (Số ví dụ: 125598)

Nhiệm vụ 1. (20% trên 100) 1)
Triển khai một sản phẩm phần mềm bằng Python bằng cách sử dụng các biểu thức chính quy theo tùy chọn được trình bày trong bảng.
2) Đưa ra ít nhất 5 bài kiểm tra cho chương trình của bạn. Mỗi bài kiểm tra là một thực thể riêng biệt được chuyển đến biểu thức chính quy để xử lý. Đối với mỗi bài kiểm tra, bạn phải độc lập (không sử dụng biểu thức chính quy) để tìm câu trả lời đúng. Sau đó so sánh đáp án mà chương trình đưa ra và đáp án mà chính bạn nhận được. 3) Chương trình phải đếm số lượng biểu tượng cảm xúc thuộc một loại nhất định (loại biểu tượng cảm xúc được mô tả trong bảng tùy chọn) trong văn bản đề xuất. Tất cả các biểu tượng cảm xúc đều có cấu trúc sau:

[mắt][mũi][miệng].
Một lựa chọn là các bộ mắt, mũi và miệng khác nhau.

Номер в ИСУ % 5	Глаза	Номер в ИСУ % 4	Нос	Номер в ИСУ % 7	Рот
0	:	0	-	0	(
1	;	1	<	1)
2	X	2	-{	2	O
3	8	3	<{	3	
4	=			4	\
				5	/
				6	P

Biểu tượng cảm xúc mẫu: 8<{P

Nhiệm vụ 2. (40% trên 100)

1) Triển khai một sản phẩm phần mềm bằng Python bằng cách sử dụng các biểu thức chính quy theo tùy chọn được trình bày trong bảng. 2) Đưa ra ít nhất 5 bài kiểm tra cho chương trình của bạn. 3) Kiểm tra chương trình của bạn bằng các bài kiểm tra này.

Số trong MIS % 6	Bài tập								
0	<p>Viết một biểu thức chính quy để kiểm tra tính chính xác của email và trả về máy chủ thư dưới dạng phản hồi (máy chủ thư là một phần của email đứng sau "@").</p> <p>Để đơn giản, chúng ta sẽ giả sử rằng địa chỉ bưu điện có thể chứa các chữ cái, số, "." và "-", còn máy chủ thư chỉ có các chữ cái và ".". Trong trường hợp này, máy chủ thư phải chứa cấp cao nhất của tên miền (".ru", ".com", v.v.)</p> <p>Ví dụ:</p> <table><tr><td>Đầu ra đầu vào</td><td></td></tr><tr><td>sinh viên.spam@yandex.ru yandex.ru</td><td></td></tr><tr><td>example@example Thất bại!</td><td></td></tr><tr><td>example@example.com example.com</td><td></td></tr></table>	Đầu ra đầu vào		sinh viên.spam@yandex.ru yandex.ru		example@example Thất bại!		example@example.com example.com	
Đầu ra đầu vào									
sinh viên.spam@yandex.ru yandex.ru									
example@example Thất bại!									
example@example.com example.com									
1	<p>Sinh viên Vasya thực sự thích khóa học Bảo mật máy tính. Một ngày nọ, Vasya được giao bài tập về nhà để mã hóa dữ liệu được gửi trong tin nhắn. Không cần suy nghĩ kỹ, Vasya quyết định thay thế tất cả các số nguyên bằng hàm của số này. Ông đã nghĩ ra một hàm đơn giản: $3 \cdot 2 + 5$, trong đó 3 là số ban đầu. Giúp Vasya làm bài tập về nhà.</p> <p>Ví dụ:</p> <table><tr><td>Nhập</td><td>Phản kết luận</td></tr><tr><td>$20 + 22 = 42$</td><td>$1205 + 1457 = 5297$</td></tr></table>	Nhập	Phản kết luận	$20 + 22 = 42$	$1205 + 1457 = 5297$				
Nhập	Phản kết luận								
$20 + 22 = 42$	$1205 + 1457 = 5297$								
2	<p>Họ đã đăng danh sách những người có học bổng cho học kỳ hiện tại, bao gồm danh sách những người, tên đầy đủ và mã số nhóm của người này. Bạn quyết định chơi khăm một số bạn cùng lớp và xóa họ khỏi danh sách. Sử dụng biểu thức chính quy, tìm tất cả học sinh trong nhóm của bạn có tên viết tắt bắt đầu bằng cùng một chữ cái và loại bỏ chúng khỏi danh sách.</p> <p>Ví dụ (nhóm P000):</p> <table><tr><td>Đầu</td><td>Phản kết luận</td></tr><tr><td>vào Petrov P.P. P000 Anishchenko A.A. P33113</td><td>Anishchenko A.A. P33113 Ví dụ E.V. P000</td></tr></table>	Đầu	Phản kết luận	vào Petrov P.P. P000 Anishchenko A.A. P33113	Anishchenko A.A. P33113 Ví dụ E.V. P000				
Đầu	Phản kết luận								
vào Petrov P.P. P000 Anishchenko A.A. P33113	Anishchenko A.A. P33113 Ví dụ E.V. P000								

	<div>Ví dụ E.V. P000 Ivanov I.I. P000</div>					
3	<div>Văn bản được đưa ra. Bạn cần tìm bất kỳ đoạn nào trong đó có từ “VT” xuất hiện đầu tiên, sau đó không quá 4 từ và sau đó là từ “ITMO” xuất hiện. Để đơn giản, chúng ta sẽ coi bất kỳ chuỗi chữ cái, số và dấu “_” nào (nghĩa là các ký tự \w) là một từ. Ví dụ:</div> <table><tr><td>Đầu</td><td>Phản kết luận</td></tr><tr><td>vào Bạn có biết PI&CT là tốt nhất giảng viên tại ITMO?</td><td>PI&CT là bộ phận tốt nhất tại ITMO</td></tr></table>	Đầu	Phản kết luận	vào Bạn có biết PI&CT là tốt nhất giảng viên tại ITMO?	PI&CT là bộ phận tốt nhất tại ITMO	
Đầu	Phản kết luận					
vào Bạn có biết PI&CT là tốt nhất giảng viên tại ITMO?	PI&CT là bộ phận tốt nhất tại ITMO					
4	<div>Văn bản được đưa ra. Bạn cần tìm tất cả các tên trong văn bản, sắp xếp chúng theo thứ tự bảng chữ cái. Để đơn giản, chúng ta coi họ là một từ có chữ in hoa theo sau là chữ cái đầu. Ví dụ:</div> <table><tr><td>Sinh</td><td>Phản kết luận</td></tr><tr><td>viên đầu vào Vasya nhớ lại điều đó trong bài giảng của mình Boldyreva E.A. đã đề cập đến những sinh viên cuối cấp sẽ giúp đỡ cô: Anishchenko A.A. và Máy E.A.</td><td>Anishchenko Boldyreva Xe hơi</td></tr></table>	Sinh	Phản kết luận	viên đầu vào Vasya nhớ lại điều đó trong bài giảng của mình Boldyreva E.A. đã đề cập đến những sinh viên cuối cấp sẽ giúp đỡ cô: Anishchenko A.A. và Máy E.A.	Anishchenko Boldyreva Xe hơi	
Sinh	Phản kết luận					
viên đầu vào Vasya nhớ lại điều đó trong bài giảng của mình Boldyreva E.A. đã đề cập đến những sinh viên cuối cấp sẽ giúp đỡ cô: Anishchenko A.A. và Máy E.A.	Anishchenko Boldyreva Xe hơi					
5	<div>Anatoly đã đăng một bài viết kèm theo lịch trình bổ sung. các lớp học về khoa học máy tính, nhưng lại trộn lẫn thời gian ở khắp mọi nơi. Vì vậy, bạn cần thay thế tất cả các lần xuất hiện thời gian bằng một chuỗi (TBD). Thời gian là một chuỗi có dạng HH:MM:SS hoặc HH:MM, trong đó HH là một số từ 00 đến 23, MM và SS là các số từ 00 đến 59. Ví dụ: Đầu</div> <table><tr><td>vào</td><td>Phản kết luận</td></tr><tr><td>Các em học sinh thân mến! Dịch vụ bổ sung được lên kế hoạch vào Thứ Bảy tuần này lúc 15:00. bài học trong 2 giờ. Tức là vào lúc 17:00:01 chắc chắn sẽ kết thúc.</td><td>Các em học sinh thân mến! Thứ bảy này tại (TBD) một sự kiện bổ sung đã được lên kế hoạch. bài học trong 2 giờ. Đó là trong (TBD) nó chắc chắn sẽ kết thúc.</td></tr></table>	vào	Phản kết luận	Các em học sinh thân mến! Dịch vụ bổ sung được lên kế hoạch vào Thứ Bảy tuần này lúc 15:00. bài học trong 2 giờ. Tức là vào lúc 17:00:01 chắc chắn sẽ kết thúc.	Các em học sinh thân mến! Thứ bảy này tại (TBD) một sự kiện bổ sung đã được lên kế hoạch. bài học trong 2 giờ. Đó là trong (TBD) nó chắc chắn sẽ kết thúc.	
vào	Phản kết luận					
Các em học sinh thân mến! Dịch vụ bổ sung được lên kế hoạch vào Thứ Bảy tuần này lúc 15:00. bài học trong 2 giờ. Tức là vào lúc 17:00:01 chắc chắn sẽ kết thúc.	Các em học sinh thân mến! Thứ bảy này tại (TBD) một sự kiện bổ sung đã được lên kế hoạch. bài học trong 2 giờ. Đó là trong (TBD) nó chắc chắn sẽ kết thúc.					

Nhiệm vụ 3. (40% trên 100)

1. Xác định số tùy chọn là phần còn lại của việc chia số trong ISU cho 36. Nếu không có lớp học nào vào một ngày nhất định trong tuần thì hãy tăng số tùy chọn lúc tám giờ.

2. Nghiên cứu dạng Backus-Naur.

3. Nghiên cứu đặc điểm các giao thức và định dạng trao đổi thông tin giữa các hệ thống: JSON, YAML, XML. 4.

Hiểu cấu trúc của trang với lịch trình cho nhóm của bạn: <https://itmo.ru/ru/schedule/0/P3110/schedule.htm>

5. Dựa trên cấu trúc của lịch trình cho một ngày cụ thể, hãy tạo một tệp có lịch trình ở định dạng được chỉ định trong nhiệm vụ làm lịch trình

ban đầu. 6. Nhiệm vụ bắt buộc: viết chương trình bằng Python 3.x để phân tích và chuyển đổi tệp nguồn thành tệp mới. 7. Bạn không thể sử dụng các thư viện có sẵn, bao gồm các biểu thức chính quy trong Python và các thư viện để tải tệp XML.

8. Nhiệm vụ bổ sung số 1 (cho phép bạn đạt +10 phần trăm số điểm BarS tối đa cho phòng thí nghiệm này). a) Tìm các thư viện làm sẵn có chức năng phân tích cú pháp và chuyển đổi tệp tương tự. b) Viết lại mã nguồn bằng cách sử dụng các thư

viện tìm được. Biểu thức thông thường cũng không thể được sử dụng. c) So sánh các kết quả thu được và giải thích sự

giống/khác nhau của chúng. 9. Nhiệm vụ bổ sung số 2 (cho phép bạn đạt

+10 phần trăm số điểm BarS tối đa cho phòng thí nghiệm này). a) Viết lại mã nguồn để bao gồm việc sử dụng các biểu thức thông thường. b)

So sánh kết quả thu được và giải thích sự giống/khác nhau.

Các lựa chọn cho nhiệm vụ 3:

Tùy chọn số.	Định dạng gốc	Định dạng kết quả	Ngày trong tuần
0	YAML	XML	Thứ hai
1	JSON	XML	Thứ hai
2	XML	JSON	Thứ hai
3	JSON	YAML	Thứ hai
4	YAML	JSON	Thứ hai
5	XML	YAML	Thứ hai
6	YAML	XML	Thứ ba
7	JSON	XML	Thứ ba
8	XML	JSON	Thứ ba
9	JSON	YAML	Thứ ba
10	YAML	JSON	Thứ ba
11	XML	YAML	Thứ ba
12	YAML	XML	Thứ Tư
13	JSON	XML	Thứ Tư
14	XML	JSON	Thứ Tư
15	JSON	YAML	Thứ Tư
16	YAML	JSON	Thứ Tư
17	XML	YAML	Thứ Tư
18	YAML	XML	Thứ năm
19	JSON	XML	Thứ năm
20	XML	JSON	Thứ năm
21	JSON	YAML	Thứ năm
22	YAML	JSON	Thứ năm
23	XML	YAML	Thứ năm
24	YAML	XML	Thứ sáu
25	JSON	XML	Thứ sáu
26	XML	JSON	Thứ sáu
27	JSON	YAML	Thứ sáu
28	YAML	JSON	Thứ sáu
29	XML	YAML	Thứ sáu
30	YAML	XML	Thứ bảy
31	JSON	XML	Thứ bảy
32	XML	JSON	Thứ bảy
33	JSON	YAML	Thứ bảy
34	YAML	JSON	Thứ bảy
35	XML	YAML	Thứ bảy

Yêu cầu và bố cục của báo cáo

1. Báo cáo phải được hoàn thành trên khổ giấy A4, sử dụng Microsoft Word, Libre Office, v.v.

2. Báo cáo phải bắt đầu bằng trang tiêu đề có tên trường đại học và khoa, số và tên công việc trong phòng thí nghiệm, môn học, tên sinh viên, mã nhóm, tên giáo viên, thành phố và năm.

3. Báo cáo phải có mục lục được tập hợp tự động (các phần bắt buộc - Nhiệm vụ, Các giai đoạn thực hiện chính, Kết luận, Danh sách nguồn sử dụng).

4. Báo cáo phải có các hình ảnh được thiết kế và ký tên theo GOST 7.32-2017 "Báo cáo về công việc nghiên cứu. Cấu trúc và quy tắc thiết kế" (ít nhất một hình ảnh) và danh sách các tài liệu tham khảo có liên kết đến các nguồn (ít nhất hai nguồn).

5. Các trang của báo cáo phải được đánh số, không đánh số trang đầu đề.

6. Báo cáo phải cung cấp ngắn gọn mô tả về sự cố đang được giải quyết, danh sách đầy đủ các chương trình .py và nội dung của tệp ở định dạng nguồn và kết quả.

7. Cung cấp báo cáo dưới dạng điện tử. Theo yêu cầu của giáo viên, bạn phải sẵn sàng biên dịch và chạy chương trình của mình.

Chuẩn bị bảo chữa 1.

Nghiên cứu và củng cố những tài liệu cần thiết từ cuốn sách sau:

Lyamin A.V., Cherepovskaya E.N. Lập trình hướng đối tượng.

Xưởng máy tính. - St. Petersburg: Đại học ITMO, 2017. - 143 tr. - Chế độ truy cập: <https://books.ifmo.ru/file/pdf/2256.pdf>.

2. Đọc và lặp lại thông tin từ bài viết trên Wikipedia: https://ru.wikipedia.org/wiki/Backus_Form_-_Naur.

3. Đọc và lặp lại thông tin trong

bài viết "Viết một trình phân tích cú pháp tĩnh tế bằng Python": <https://habr.com/ru/post/309242/>.

4. Có thể giải

thích từng dòng chương trình được trình bày trong báo cáo.

5. Khi bảo vệ báo cáo, bạn phải trả lời được các câu hỏi về hoạt động của chương trình, các câu hỏi dựa trên tài liệu của bài giảng số 3 và số 4 và các câu

hỏi sau: 1) Markup và Markdown khác nhau như

thế nào? 2) PROTOBUF có gì đặc biệt so với các định dạng khác? 3) Định dạng CSV

khác với định dạng TSV như thế nào?

4) Lý do cho sự xuất hiện liên tục của các định dạng trình bày dữ liệu mới là gì?

5) Các ký tự '>' và '<' được thể hiện như thế nào trong XML? 6) Tuần tự hóa dữ liệu là gì? 7) Các nhận

xét được chỉ định trong YAML như thế nào? 8) Giải thích cách tạo tiêu đề ở các cấp độ khác nhau trong ngôn ngữ đánh dấu Markdown, định dạng mã, hiển thị in đậm, in nghiêng và gạch ngang chữ?

9) Những định dạng trao đổi dữ liệu nào được sử dụng trong các ứng dụng nhắn tin tức thời phổ biến hiện đại (Viber, WhatsApp,

Telegram, v.v.)? 10) Chữ viết tắt SVG có nghĩa

là gì? 11) Cho ví dụ về cách sử dụng thẻ trong HTML để tạo siêu liên kết đến url. 12) Văn

bản JSON có thể biểu thị hai cấu trúc nào ở dạng được mã hóa?