

COS20015 Fundamentals of Data Management

Workshop 1 & 2 completed tasks

Student Name 104268899

Workshop 1

Part 1:

REGULAR EXPRESSION

22 matches (126 steps, 1.33ms)

TEST STRING

Meeting-Pre-print-CDROM.

Byon, Y., J., Abdulhai, B., & Shalaby, A. S. (2007). Impact of sampling rate of GPS-enabled cell phones on mode detection and GIS map matching performance. In Transportation Research Board 86th Annual Meeting (No. 071795).

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies, 17, 285-297.

Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. Transportation Research Part A: Policy and Practice, 44, 830-840.

Deng, Z., & Ji, M. (2010, August). Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. In Traffic and Transportation Studies 2010 (pp. 768-777). ASCE.

Gonzalez, P., Weinstein, J., Barbeau, S., Labrador, M., Winters, P., Georggi, N., & Perez, R. (2008, November). Automating mode detection using neural networks and assisted gps data collected using gps-enabled mobile phones. In 15th World Congress on Intelligent Transportation Systems.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems, 36(2), 131-139.

Hato, E., Shinji, I., & Mitani, T. (2006). Development of MoALs (Mobile Activity Loggers) supported by GPS phones for Travel Behavior Analysis. Proceedings of the 85th Annual Meeting of the Transportation Research Board, January 2006, Washington D.C.

Itsubo, S., & Hato, E. (2006). A study of the effectiveness of a household travel survey using GPS.

EXPLANATION

\\(matches the character ((with index 40₁₀ (28₁₆ or 50₈) literally (case sensitive)

\\d matches a digit (equivalent to [0-9])

{4} matches the previous token exactly 4 times

\\) matches the character) with index 41₁₀ (29₁₆ or 51₈) literally (case sensitive)

Global pattern flags

g modifier: global. All matches (don't return after first match)

m modifier: multi line. Causes ^ and \$ to match the begin/end of each line (not only begin/end of string)

MATCH INFORMATION

Match	Start	End	Text
Match 1	288	294	(2007)
Match 2	491	497	(2009)
Match 3	761	767	(2010)
Match 4	1557	1563	(2012)
Match 5	1721	1727	(2006)

QUICK REFERENCE

Search reference

All Tokens

Common Token...

General Tokens

anchors

Meta Sequences

Quantifiers

Group Constructs

A single character of: a, b or c [abc]

A character except: a, b or c [^abc]

A character in the range: a-z [a-z]

A character not in the range: a-z [^a-z]

A character in the range: a-z or A-Z [a-zA-Z]

Any single character .

Alternate - match either a or b a|b

Any whitespace character \s

Solution:
\\(\\d{4}\\)

Part 2:

REGULAR EXPRESSION

20 matches (346 steps, 765µs)

TEST STRING

References

Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M., & Samaga, U. (2004, January). Eighty-weeks of gps traces, approaches to enriching trip information. In Transportation Research Board 83rd Annual Meeting-Pre-print-CDROM.

Byon, Y., J., Abdulhai, B., & Shalaby, A. S. (2007). Impact of sampling rate of GPS-enabled cell phones on mode detection and GIS map matching performance. In Transportation Research Board 86th Annual Meeting (No. 071795).

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies, 17, 285-297.

Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. Transportation Research Part A: Policy and Practice, 44, 830-840.

Deng, Z., & Ji, M. (2010, August). Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. In Traffic and Transportation Studies 2010 (pp. 768-777). ASCE.

Gonzalez, P., Weinstein, J., Barbeau, S., Labrador, M., Winters, P., Georggi, N., & Perez, R. (2008, November). Automating mode detection using neural networks and assisted gps data collected using gps-enabled mobile phones. In 15th World Congress on Intelligent Transportation Systems.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems, 36(2), 131-139.

Hato, E., Shinji, I., & Mitani, T. (2006). Development of MoALs (Mobile Activity Loggers) supported by GPS phones for Travel Behavior Analysis. Proceedings of the 85th Annual Meeting of the Transportation Research Board, January 2006, Washington D.C.

EXPLANATION

\\(matches the character ((with index 40₁₀ (28₁₆ or 50₈) literally (case sensitive)

\\(200[0-9]|201[0-9]|202[0-2]\\) matches a single character present in the list below [0-9]

1st Alternative 200[0-9]

200 matches the characters 200 literally (case sensitive)

Match a single character present in the list below [0-9]

2nd Alternative 201[0-9]

201 matches the characters 201 literally (case sensitive)

Match a single character present in the list below [0-9]

MATCH INFORMATION

Match	Start	End	Text
Match 1	288	294	(2007)
Match 2	491	497	(2009)
Match 3	761	767	(2010)

QUICK REFERENCE

Search reference

All Tokens

Common Token...

General Tokens

anchors

Meta Sequences

Quantifiers

Group Constructs

A single character of: a, b or c [abc]

A character except: a, b or c [^abc]

A character in the range: a-z [a-z]

A character not in the range: a-z [^a-z]

A character in the range: a-z or A-Z [a-zA-Z]

Any single character .

Alternate - match either a or b a|b

Any whitespace character \s

Solution: \\((200[0-9]|201[0-9]|202[0-2])\\)

Part 3:

REGULAR EXPRESSION

8 matches (520 steps, 425µs)

TEST STRING

References

Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M., & Samaga, U. (2004, January). Eighty weeks of gps traces, approaches to enriching trip information. In Transportation Research Board 83rd Annual Meeting Pre-print CDROM.

Byon, Y. J., Abdulhai, B., & Shalaby, A. S. (2007). Impact of sampling rate of GPS-enabled cell phones on mode detection and GIS map matching performance. In Transportation Research Board 86th Annual Meeting (No. 071795).

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies, 17, 285-297.

Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. Transportation Research Part A: Policy and Practice, 44, 830-840.

Deng, Z., & Ji, M. (2010, August). Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. In Traffic and Transportation Studies 2010 (pp. 768-777). ASCE.

Gonzalez, P., Weinstein, J., Barbeau, S., Labrador, M., Winters, P., Georggi, N. L., & Perez, R. (2008, November). Automating mode detection using neural networks and assisted gps data collected using gps-enabled mobile phones. In 15th World Congress on Intelligent Transportation Systems.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems, 36(2), 131-139.

EXPLANATION

/ \((200[0-9]|201[0-9]|202[0-2]), (January|February|March|April|May|June|July|August|September|October|November|December)\) / gm

1st Capturing Group (200[0-9]|201[0-9]|202[0-2])

1st Alternative 200[0-9]

200 matches the characters 200 literally (case sensitive)

Match a single character present in the list below [0-9]

0-9 matches a single character in the range between 0 (index 48) and 9 (index 57) (case sensitive)

2nd Alternative 201[0-9]

201 matches the characters 201 literally (case sensitive)

2nd Alternative 202[0-2]

202 matches the characters 202 literally (case sensitive)

Match a single character present in the list below [0-2]

0-2 matches a single character in the range between 0 (index 48) and 2 (index 50) (case sensitive)

MATCH INFORMATION

Match	Start	End	Text
Match 1	84	99	(2004, January)
Group 1	85	89	2004
Group 2	91	98	January
Match 2	1005	1019	(2010, August)
Group 1	1006	1010	2010
Group 2	1012	1018	August

QUICK REFERENCE

Search reference

All Tokens

Common Token

General Tokens

Anchor

Meta Sequences

Quantifiers

Group Constructs

A single character of: a, b or c [abc]

A character except: a, b or c [^abc]

A character in the range: a-z [a-z]

A character not in the range: a-z [^a-z]

A character in the range: a-z or A-Z [a-zA-Z]

Any single character .

Alternate - match either a or b a|b

Any whitespace character \s

Solution: `\((200[0-9]|201[0-9]|202[0-2]), (January|February|March|April|May|June|July|August|September|October|November|December)\)`

Part 4:

REGULAR EXPRESSION

18 matches (2709 steps, 565µs)

TEST STRING

References

Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M., & Samaga, U. (2004, January). Eighty weeks of gps traces, approaches to enriching trip information. In Transportation Research Board 83rd Annual Meeting Pre-print CDROM.

Byon, Y. J., Abdulhai, B., & Shalaby, A. S. (2007). Impact of sampling rate of GPS-enabled cell phones on mode detection and GIS map matching performance. In Transportation Research Board 86th Annual Meeting (No. 071795).

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies, 17, 285-297.

Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. Transportation Research Part A: Policy and Practice, 44, 830-840.

Deng, Z., & Ji, M. (2010, August). Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. In Traffic and Transportation Studies 2010 (pp. 768-777). ASCE.

Gonzalez, P., Weinstein, J., Barbeau, S., Labrador, M., Winters, P., Georggi, N. L., & Perez, R. (2008, November). Automating mode detection using neural networks and assisted gps data collected using gps-enabled mobile phones. In 15th World Congress on Intelligent Transportation Systems.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems, 36(2), 131-139.

Hato, F., Shinji, T., & Mitani, T. (2006). Development of MoAIs (Mobile Activity Lappers supported by

EXPLANATION

/ \b[A-Z][a-zA-Z]+, [A-Z]\. (?:[A-Z])? \b / gm

\b assert position at a word boundary

Match a single character present in the list below [A-Z]

A-Z matches a single character in the range between A (index 65) and Z (index 90) (case sensitive)

Match a single character present in the list below [a-zA-Z]

a matches the previous token between one and unlimited times, as many times as possible, giving back as needed (greedy)

a-z matches a single character in the range between a (index 97) and z (index 122) (case sensitive)

Match a single character present in the list below [A-Z]

A matches a single character in the range between A (index 65) and Z (index 90) (case sensitive)

MATCH INFORMATION

Match	Start	End	Text
Match 1	13	27	Axhausen, K. W.
Match 2	244	254	Byon, Y. J.
Match 3	273	286	Shalaby, A. S.
Match 4	1280	1293	Georggi, N. L.
Match 5	1543	1555	Lawson, C. T.

QUICK REFERENCE

Search reference

All Tokens

Common Token

General Tokens

Anchor

Meta Sequences

Quantifiers

Group Constructs

A single character of: a, b or c [abc]

A character except: a, b or c [^abc]

A character in the range: a-z [a-z]

A character not in the range: a-z [^a-z]

A character in the range: a-z or A-Z [a-zA-Z]

Any single character .

Alternate - match either a or b a|b

Any whitespace character \s

Solution: `\b[A-Z][a-zA-Z]+, [A-Z]\. (?:[A-Z])? \b`

Workshop 2

Part 1:

Liberation Sans 10 pt **B** *I* U **A**

D10 **fx** Σ =

	B	C	D	E	F	G	H
1	ClientName	PhysioName	Date	Time	Duration	Condition	
2	Sam Huynh	Bruno	17/04/23	09:00:00	30 minutes	broken wrist	
3	Andrea Kostyanska	Sylvia	18/04/23	15:30:00	30 minutes	lumbar spine	
4	Philip Chen	Bruno	18/04/23	22:30:00	40 minutes	neck	
5							
6							
7							
8							
9							
10							
11							

Part 2:

[illegible]

Part 3:

	A	B	C	D	E	F	G
1	Appointment ID – INTEGER, PRIMARY KEY	ClientName – VARCHAR(25)	PhysioName – VARCHAR(15)	Date – DATE	Time – TIME	DurationMinutes – INTEGER	Condition – VARCHAR(30)
2		1 Sam Huynh	Bruno	17/04/23	09:00:00	30	broken wrist
3		2 Andrea Kostyanska	Sylvia	18/04/23	15:30:00	30	lumbar spine
4		3 Philip Chen	Bruno	18/04/23	22:30:00	40	neck
5							
6							
7	<p>We have to use AppointmentID as the primary key for this table because it ensures that each appointment is uniquely identifiable. However, if AppointmentID is not available, we can use a composite key made up of ClientName, Date, and Time.</p>						
8							