

Sequential plugins

Strings, Chains, Sequences, Intervals

The **GALACTIC** Organization <contact@thegalactic.org>



2019-2022 

¹© 2019-2022 the **GALACTIC** Organization. This document is licensed under CC-by-nc-nd
(<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>)



Sequences

Sequences are built from a set of elements of the same type called **Alphabet**

For example, DNA is composed of a succession of nucleotides. There are four different nucleotides that constitute the alphabet,

$$\Sigma = \{\textit{adenosine (A)}, \textit{cytidine (C)}, \textit{guanosine (G)}, \textit{thymidine (T)}\}.$$

2022-02-03

Sequential plugins
└ Introduction
└ Sequences
└ Sequences

Sequences

Sequences are built from a set of elements of the same type called **Alphabet**
For example, DNA is composed of a succession of nucleotides. There are four different nucleotides that constitute the alphabet.
 $\Sigma = \{\textit{adenosine (A)}, \textit{cytidine (C)}, \textit{guanosine (G)}, \textit{thymidine (T)}\}.$

Sequential plugins

Sequences

Sequential plugins are developed for the **GALACTIC** framework to mine sequential data.

We distinguish 3 types of plugins.

Characteristics



Descriptions



Strategies



2022-02-03

Sequential plugins

- └ Introduction
 - └ Sequential plugins
 - └ Sequential plugins

Sequential plugins

Sequences

Sequential plugins are developed for the GALACTIC framework to mine sequential data. We distinguish 3 types of plugins.

Characteristics

Descriptions

Strategies



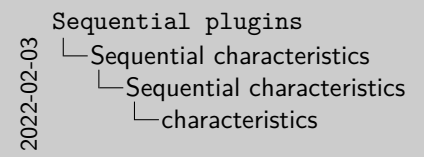
characteristics

Characteristics

Characteristics define the types of data taken for analysis. For example, binary data, numeric, sequential ... etc.

We have implemented four plugins that represent four types of sequential data:

Strings	$a_1 a_2 \dots a_n$
Chains (sequences)	$[a_1, a_2, \dots, a_n]$
Temporal sequences	$[(t_1, a_1), (t_2, a_2), \dots, (t_n, a_n)]$
Interval sequences	$[((\underline{t}_1, \bar{t}_1), a_1), ((\underline{t}_2, \bar{t}_2), a_2), \dots, ((\underline{t}_n, \bar{t}_n), a_n)]$



characteristics

Characteristics
Characteristics define the types of data taken for analysis. For example, binary data, numeric, sequential ... etc.

We have implemented four plugins that represent four types of sequential data:

Strings	$a_1 a_2 \dots a_n$
Chains (sequences)	$[a_1, a_2, \dots, a_n]$
Temporal sequences	$[(t_1, a_1), (t_2, a_2), \dots, (t_n, a_n)]$
Interval sequences	$[((\underline{t}_1, \bar{t}_1), a_1), ((\underline{t}_2, \bar{t}_2), a_2), \dots, ((\underline{t}_n, \bar{t}_n), a_n)]$

Descriptions

A description δ is an application which provides predicates describing a set of objects $A \subseteq G$ according to their characteristics.

Depending on the characteristics, We have implemented some descriptions that use predicates of several types "*subsequence of*", "*super-sequence of*", etc.

Descriptions	Predicates
Simple	Maximal common K-subsequences
Complete	Maximal common subsequences
Affix	Maximal subsequences: prefix-suffix
Distance	Maximal subsequences with distances
Time Frame	Maximal subsequences and minimal supersequences

2022-02-03

Sequential plugins
 └ Sequential descriptions
 └ Sequential descriptions

Descriptions
 A description δ is an application which provides predicates describing a set of objects $A \subseteq G$ according to their characteristics.

Depending on the characteristics, We have implemented some descriptions that use predicates of several types "*subsequence of*", "*super-sequence of*", etc.

Descriptions	Predicates
Simple	Maximal common K-subsequences
Complete	Maximal common subsequences
Affix	Maximal subsequences: prefix-suffix
Distance	Maximal subsequences with distances
Time Frame	Maximal subsequences and minimal supersequences

Strategies

A strategy σ refines a concept $(A, \delta(A))$ into subconcepts composed of a reduced set of objects $A' \subset A$ described by bigger subsequences $\delta(A) \sqsubseteq \delta(A')$. It corresponds to the generation of immediate predecessors in the lattice.

- ▶ Simple: with generation of all possible sub-sequences.
- ▶ Augmented: with generation of subsequences by adding an item of the alphabet to the predicates of the description.
- ▶ Distance: based on the distances between elements of the sequences.
- ▶ Alphabet: for the description of super-sequences; an element of the alphabet is removed each time.

2022-02-03

Sequential plugins
└ Sequential strategies
└ Sequential strategies

Strategies

A strategy σ refines a concept $(A, \delta(A))$ into subconcepts composed of a reduced set of objects $A' \subset A$ described by bigger subsequences $\delta(A) \sqsubseteq \delta(A')$. It corresponds to the generation of immediate predecessors in the lattice.

- ▶ Simple: with generation of all possible sub-sequences.
- ▶ Augmented: with generation of subsequences by adding an item of the alphabet to the predicates of the description.
- ▶ Distance: based on the distances between elements of the sequences.
- ▶ Alphabet: for the description of super-sequences; an element of the alphabet is removed each time.

Wine City

- ▶ This dataset is coming from the Wine City museum
- ▶ Gathered from the visits on a period of one year (May 2016 to May 2017).
- ▶ This dataset has been cleaned and processed before.
- ▶ Visitors navigate from modules to modules exploring the museum.
- ▶ The museum is open, and they are not “Guided”.

2022-02-03

Sequential plugins
└─ Examples
 └─ Wine City
 └─ Wine City

Wine City

- ▶ This dataset is coming from the Wine City museum
- ▶ Gathered from the visits on a period of one year (May 2016 to May 2017).
- ▶ This dataset has been cleaned and processed before.
- ▶ Visitors navigate from modules to modules exploring the museum.
- ▶ The museum is open, and they are not “Guided”.

Wine City

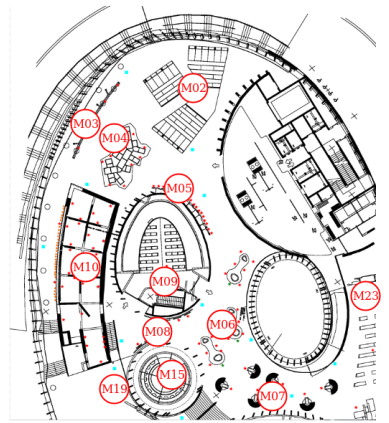


Figure 1: Modules location in the museum

2022-02-03
Sequential plugins
└─ Examples
 └─ Wine City
 └─ Wine City

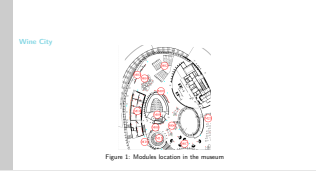
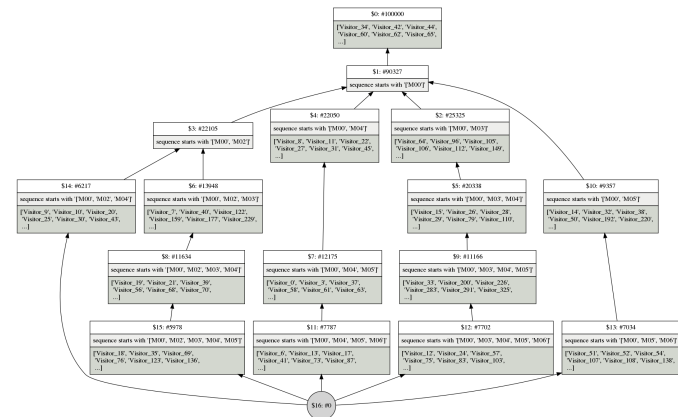


Figure 1: Modules location in the museum

Wine City with prefix match

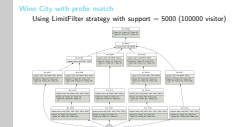
Using LimitFilter strategy with support = 5000 (100000 visitor)



2022-02-03

Sequential plugins

- └ Examples
 - └ Wine City
 - └ Wine City with prefix match



Wine City with prefix match

\$12: #7702
sequence starts with ['M00', 'M03', 'M04', 'M05', 'M06']
['Visitor_12', 'Visitor_24', 'Visitor_57', 'Visitor_75', 'Visitor_83', 'Visitor_103', ...]

7% of visitors starts with M00 M03
M04 M05 M06

2022-02-03

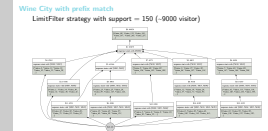
Sequential plugins

- └ Examples
 - └ Wine City
 - └ Wine City with prefix match

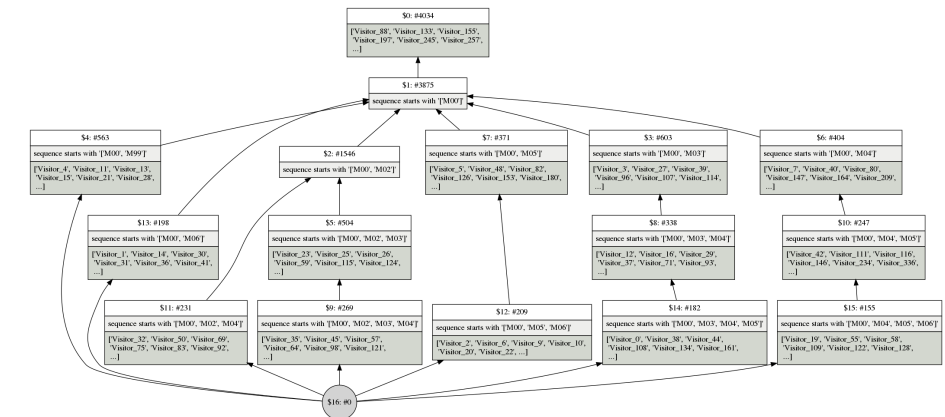
Wine City with prefix match

\$12: #7702
sequence starts with ['M00', 'M03', 'M04', 'M05', 'M06']
['Visitor_12', 'Visitor_24', 'Visitor_57', 'Visitor_75', 'Visitor_83', 'Visitor_103', ...]

7% of visitors starts with M00 M03
M04 M05 M06



Wine City with prefix match
LimitFilter strategy with support = 150 (~9000 visitor)



GéoLuciole

- ▶ GPS trajectories of people's movements in the city of La Rochelle in France.
- ▶ The data have been collected by a specific application named *GéoLuciole*, developed for the DA3T² project.
- ▶ 20 trajectories.

²System for the Analysis of Numerical Traces for the development of Tourist Territories (Dispositif d'Analyse des Traces numériques pour la valorisation des Territoires Touristiques)

2022-02-03

Sequential plugins

- └ Examples
 - └ GéoLuciole
 - └ GéoLuciole

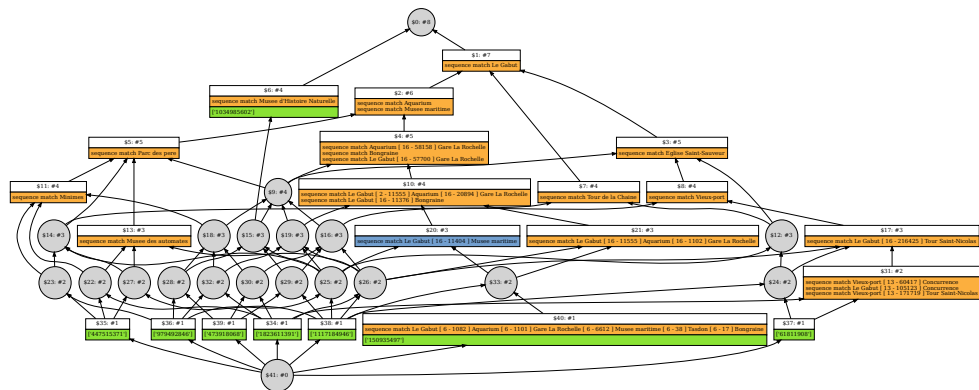
GéoLuciole

- ▶ GPS trajectories of people's movements in the city of La Rochelle in France.
- ▶ The data have been collected by a specific application named *GéoLuciole*, developed for the DA3T² project.
- ▶ 20 trajectories.

²System for the Analysis of Numerical Traces for the development of Tourist Territories (Dispositif d'Analyse des Traces numériques pour la valorisation des Territoires Touristiques)

GéoLuciole with Complete description

Middle strategy and Complete Description.

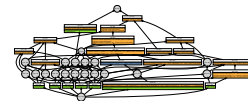


2022-02-03

Sequential plugins

- └─ Examples
 - └─ GéoLuciole
 - └─ GéoLuciole with Complete description

GéoLuciole with Complete description
Middle strategy and Complete Description.



Axes

We consider three aspects (Axis) of a trajectory : Spatial, Movement and Activities

Axis 1

["Centre-Ville", "La Guinguette", "Les Minimes", "Centre-Ville"]

Axis 2

["Stop", "Marche", "Marche", "Stop", "Stop", "Vélo", "Marche", "Stop"]

Axis 3

["Repas", "Repas", "Cinéma", "Repas", "Repos"]

2022-02-03

Sequential plugins

└─ Examples

└─ Multi-axis analysis using MCS description

└─ Axes

Axes

We consider three aspects (Axis) of a trajectory : Spatial, Movement and Activities

Axis 1

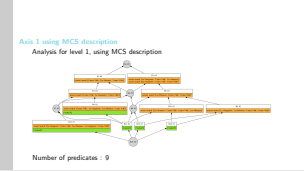
["Centre-Ville", "La Guinguette", "Les Minimes", "Centre-Ville"]

Axis 2

["Stop", "Marche", "Marche", "Stop", "Stop", "Vélo", "Marche", "Stop"]

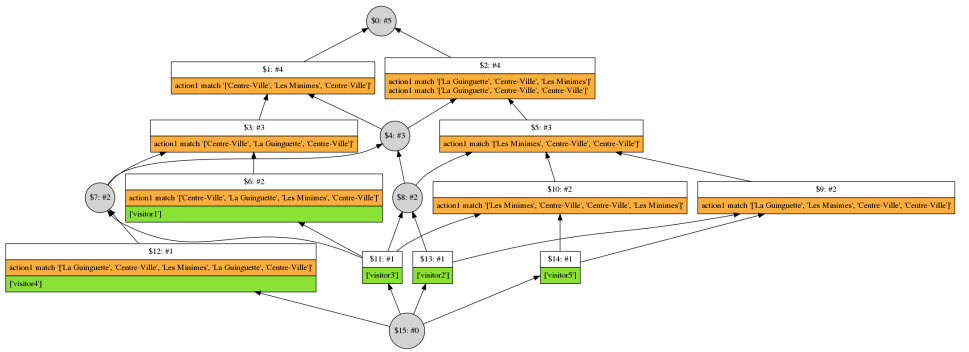
Axis 3

["Repas", "Repas", "Cinéma", "Repas", "Repos"]



Axis 1 using MCS description

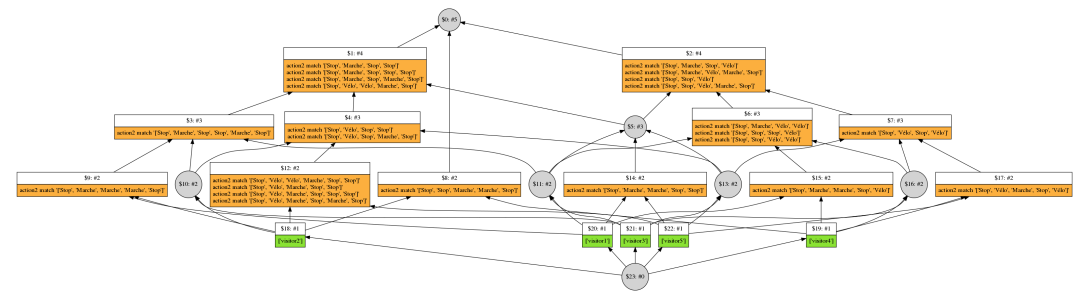
Analysis for level 1, using MCS description



Number of predicates : 9

Axis 2 using MCS description

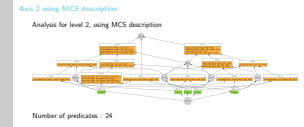
Analysis for level 2, using MCS description



Number of predicates : 24

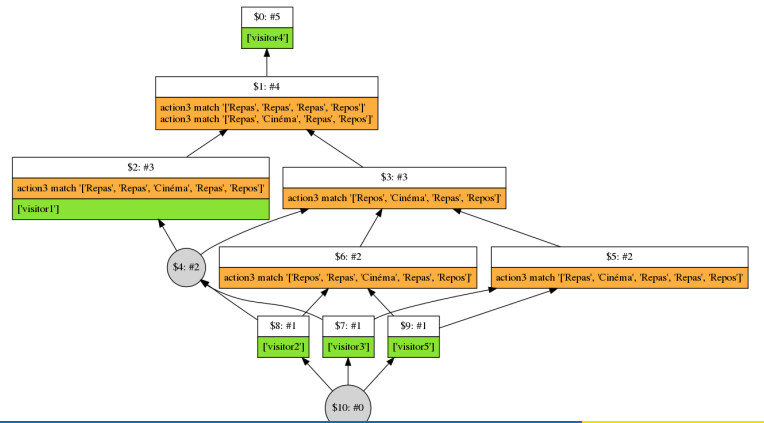
2022-02-03

- Sequential plugins
 - Examples
 - Multi-axis analysis using MCS description
 - Axis 2 using MCS description



Axis 3 using MCS description

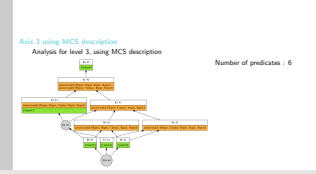
Analysis for level 3, using MCS description



Number of predicates : 6

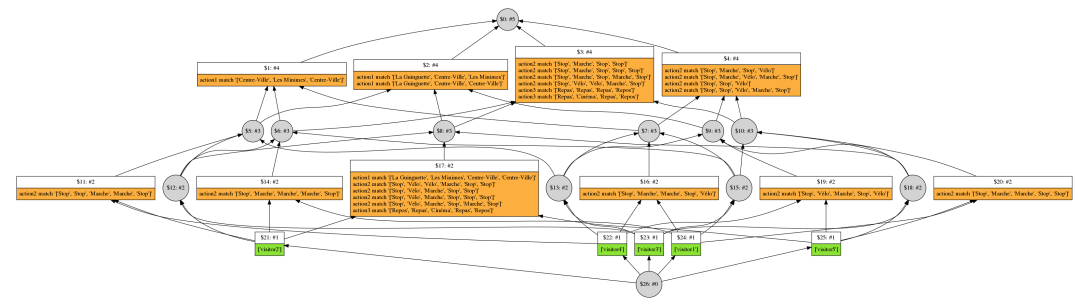
2022-02-03

- Sequential plugins
 - Examples
 - Multi-axis analysis using MCS description
 - Axis 3 using MCS description



Multi-axis analysis using MCS description

Analysis for all axis, using MCS description



Number of predicates : $24 < 39 = 9 + 24 + 6$

2022-02-03

Sequential plugins

- └ Examples
 - └ Multi-axis analysis using MCS description
 - └ Multi-axis analysis using MCS description

Multi-axis analysis using MCS description

Analysis for all axis, using MCS description



Number of predicates : $24 < 39 = 9 + 24 + 6$

Multi-axis analysis using MCS description

\$3: #4

```

action2 match ['Stop', 'Marche', 'Stop', 'Stop']
action2 match ['Stop', 'Marche', 'Stop', 'Stop', 'Stop']
action2 match ['Stop', 'Marche', 'Stop', 'Marche', 'Stop']
action2 match ['Stop', 'Vélo', 'Vélo', 'Marche', 'Stop']
action3 match ['Repas', 'Repas', 'Repas', 'Repos']
action3 match ['Repas', 'Cinéma', 'Repas', 'Repos']

```

4 individuals make these subsequences for movement (action 2):

- ▶ stop, marche, stop, stop
- ▶ stop, marche, stop, stop, stop
- ▶ stop, marche, stop, stop, stop
- ▶ stop, marche, stop, marche, stop
- ▶ stop, vélo, vélo, marche, stop

The same individuals also did these activities (action 3):

- ▶ repas, repas, repas, repos
- ▶ repas, cinéma, repas, repos

2022-02-03

Sequential plugins

└ Examples

└ Multi-axis analysis using MCS description

└ Multi-axis analysis using MCS description

Multi-axis analysis using MCS description

```

$3 #4
action2 match ['Stop', 'Marche', 'Stop', 'Stop']
action2 match ['Stop', 'Marche', 'Stop', 'Stop', 'Stop']
action2 match ['Stop', 'Marche', 'Stop', 'Marche', 'Stop']
action2 match ['Stop', 'Vélo', 'Vélo', 'Marche', 'Stop']
action3 match ['Repas', 'Repas', 'Repas', 'Repos']
action3 match ['Repas', 'Cinéma', 'Repas', 'Repos']

```

4 individuals make these subsequences for movement (action 2):

- ▶ stop, marche, stop, stop
- ▶ stop, marche, stop, stop, stop
- ▶ stop, marche, stop, marche, stop
- ▶ stop, vélo, vélo, marche, stop

The same individuals also did these activities (action 3):

- ▶ repas, repas, repas, repos
- ▶ repas, cinéma, repas, repos

Multi-axis analysis using MCS description

\$17: #2
action1 match ['La Guinguette', 'Les Minimes', 'Centre-Ville', 'Centre-Ville']
action2 match ['Stop', 'Vélo', 'Vélo', 'Marche', 'Stop', 'Stop']
action2 match ['Stop', 'Vélo', 'Marche', 'Stop', 'Stop']
action2 match ['Stop', 'Vélo', 'Marche', 'Stop', 'Stop', 'Stop']
action2 match ['Stop', 'Vélo', 'Marche', 'Stop', 'Marche', 'Stop']
action3 match ['Repos', 'Repas', 'Cinéma', 'Repas', 'Repos']

2 individuals make these subsequences for spatial (action 1):

- ▶ La Guinguette, Les minimes, Centre ville, Centre ville

For the movement (action 2):

- ▶ stop, vélo, vélo, marche, stop, stop
- ▶ stop, vélo, marche, stop, stop
- ▶ stop, vélo, marche, stop, stop, stop
- ▶ stop, vélo, marche, stop, marche, stop

And for the activities (action 3):

- ▶ repos, repas, cinéma, repas, repos

2022-02-03

Sequential plugins

└─ Examples

└─ Multi-axis analysis using MCS description

└─ Multi-axis analysis using MCS description

Multi-axis analysis using MCS description

\$17: #2
action1 match ['La Guinguette', 'Les Minimes', 'Centre-Ville', 'Centre-Ville']
action2 match ['Stop', 'Vélo', 'Vélo', 'Marche', 'Stop', 'Stop']
action2 match ['Stop', 'Vélo', 'Marche', 'Stop', 'Stop']
action2 match ['Stop', 'Vélo', 'Marche', 'Stop', 'Stop', 'Stop']
action2 match ['Stop', 'Vélo', 'Marche', 'Stop', 'Marche', 'Stop']
action3 match ['Repos', 'Repas', 'Cinéma', 'Repas', 'Repos']

2 individuals make these subsequences for spatial (action 1):

- ▶ La Guinguette, Les minimes, Centre ville, Centre ville

For the movement (action 2):

- ▶ stop, vélo, vélo, marche, stop, stop
- ▶ stop, vélo, marche, stop, stop
- ▶ stop, vélo, marche, stop, stop, stop
- ▶ stop, vélo, marche, stop, marche, stop

And for the activities (action 3):

- ▶ repos, repas, cinéma, repas, repos

Multi-axis analysis using MCS description

If we analyse the three levels separately we get:

- ▶ number of predicates: $39 = 9 + 24 + 6$
- ▶ number of concepts: $61 = 11 + 24 + 16$

With multi-axis we get:

- ▶ number of predicates: 24
- ▶ number of concepts: 27

2022-02-03

Sequential plugins

└─ Examples

└─ Multi-axis analysis using MCS description

└─ Multi-axis analysis using MCS description

Multi-axis analysis using MCS description

If we analyse the three levels separately we get:

- ▶ number of predicates: $39 = 9 + 24 + 6$
- ▶ number of concepts: $61 = 11 + 24 + 16$

With multi-axis we get:

- ▶ number of predicates: 24
- ▶ number of concepts: 27

Conclusion

We used the `NEXTPRIORITYCONCEPT` algorithm and the **GALACTIC** framework to analyse sequential data, we developed many kinds of descriptions and strategies representing different ways of analysing sequential data: strings, chains, sequences, and intervals.

2022-02-03
Sequential plugins
└─ Conclusion
 └─ Conclusion
 └─ Conclusion

Conclusion

We used the `NEXTPRIORITYCONCEPT` algorithm and the **GALACTIC** framework to analyse sequential data, we developed many kinds of descriptions and strategies representing different ways of analysing sequential data: strings, chains, sequences, and intervals.

Nicolas Jay LASH

Contact

GALACTIC web site

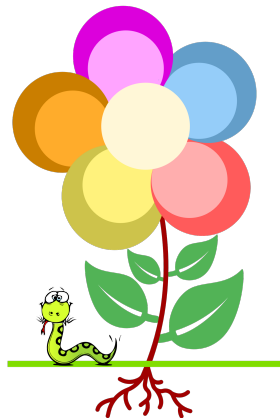
<https://galactic.univ-lr.fr>

GALACTIC mailing list

<https://ml.univ-lr.fr/sympa/info/galactic>

THANKS

For listening ;)



2022-02-03

Sequential plugins

- └ Conclusion
- └ Contact
- └ Contact

Contact

GALACTIC web site
<https://galactic.univ-lr.fr>

GALACTIC mailing list
<https://ml.univ-lr.fr/sympa/info/galactic>

THANKS

For listening ;)

