

Json exchange format syntax

Adrien Basso-Blandin

June 28, 2016

1 Json description

The Json file contains nuggets describing pieces of knowledges extracted from publications or databases. In this endeavour, it is defined as a list of nugget objects. the table 1.1 define the Json data structure.

1.1 Json detail

- A Json file is a list of nuggets.
- A nugget is a graph, it contains a list of nodes, a list of edges connecting those nodes and the list of the main nodes of this nugget (main nodes are actions). most of the time there is only one main node which is the action described by the nugget. It may also contain a comment describing it.
- A node is defined by its id and its types. It may have labels, values and upId.
- A node id is a unique identifier for each node.
- An up id is either an uniprot identifier or a unique user define identifier for nodes (for example two instance of the protein "A" must have the same upId, but two different id.
- We define four categories of nodes :
 - 'component', which contains 'agent' (protein), 'region' (may be a domain), 'keyres'(key residues) and 'flag'.
 - 'action', which contains 'bnd'(bind), 'brk'(break), 'syn'(synthetize), 'deg'(degradation), 'posMod' (positive modification (increment the value by one)), 'negMod'(negative modification, (decrement the value by one), 'mod'(change a specific value to another), 'input' (input of action nodes), 'output'(output of action nodes)
 - 'rw_rule', which contains 'set' (rewriting a subgraph into one node) and 'family' (rewriting multiple instances of a family into one node)

- 'attribute' define all the biological attributes, it may have a type as label and some values, it can be linked to all other components (except 'input' and 'output')
- A label is a string giving some humane readable informations. every nodes except 'input' and 'output' may have one. 'agent' with the same upid can't have disjoint labels.
- A value is a string defining a value that can change according to specific interactions. Only 'flag' and 'attribute' may have values. 'mod' may also have one or two corresponding to the (facultative) initial value and the final value after modification.
- An edge is linking two nodes, it has an (facultative) id if any reference to it is necessary, it also has a type and contains a reference to each nodes it is linking.
- edge must be of one of those types : 'parent' defining parenting relation between components, components and attributes or actions (or actions and attributes). 'link' define an edge between a component and an action. 'rw' define link between 'rw_rule' and other nodes. Finally, 'posInfl' and 'negInfl' define influence relations between agents, actions or agents and actions.
- An eid is a unique id for an edge.

```

Json := {"nuggets":[nugget list]}

nugget := {"comment":string,
  "nodes":[node list],
  "edges":[edge list],
  "main_nodes":[nid list]}

node := {"id":nid,
  "types":[node_type list],
  "labels":[label list],
  "values":[value list],
  "UPid":upid}

nid := 'n_' int

upid := 'up_' int | 'u_' int

node_type := "component" | "action" | "rw_rule"
  | "attribute" | "bnd" | "brk"
  | "syn" | "deg" | "posMod"
  | "negMod" | "mod" | "agent"
  | "region" | "keyres" | "flag"
  | "set" | "family" | "input" | "output"

label := string

value:= string

edge := {"id":eid,
  "type":edge_type,
  "source":nid,
  "target":nid}

edge_type := "parent" | "link" | "rw"
  | "posInfl" | "negInfl"

eid := 'e_' int

```

Table 1: Json data structure

2 example

The table 2 shows a Json example. Notice that a bnd action has 2 input, a brk has 2 output, a syn has 1 output, a deg has 1 input, a mod has 1 input and 1 output. Notice that the UPid is only here to allow the user to give what label he wants to nodes. (at the exception of agents).

```

{"nuggets": [
{
"comment": "A protein dimerisation when one is phosphorylated",
"nodes": [
{"id": 'n_0', "types": ["component", "agent"], "labels": ["A"], "UPid": 'u_0'},
{"id": 'n_1', "types": ["component", "agent"], "labels": ["A"], "UPid": 'u_0'},
{"id": 'n_2', "types": ["component", "flag"], "labels": ["phos"], "values": ['p'], "UPid": 'u_1'},
{"id": 'n_3', "types": ["action", "bnd"], "labels": ["dimerize"]},
{"id": 'n_4', "types": ["action", "input"]},
{"id": 'n_5', "types": ["action", "input"]},
{"id": 'n_6', "types": ["attribute"], "labels": ["#act_rate"], "values": [0.21]}],
"edges": [
{"type": "parent", "source": 'n_2', "target": 'n_0'},
{"type": "parent", "source": 'n_6', "target": 'n_3'},
{"type": "parent", "source": 'n_4', "target": 'n_3'},
{"type": "parent", "source": 'n_5', "target": 'n_3'},
{"type": "link", "source": 'n_0', "target": 'n_4'},
{"type": "link", "source": 'n_1', "target": 'n_5'}],
"main_nodes": ['n_3']}
], {
"comment": "A phosphorylation",
"nodes": [
{"id": 'n_7', "types": ["component", "agent"], "labels": ["A"], "UPid": 'u_0'},
{"id": 'n_8', "types": ["component", "flag"], "labels": ["phos"], "values": ['u', 'p'], "UPid": 'u_1'},
{"id": 'n_9', "types": ["action", "mod"], "labels": ["Phosphorilate"]},
{"id": 'n_10', "types": ["action", "input"]},
{"id": 'n_11', "types": ["action", "output"]},
{"id": 'n_12', "types": ["component", "agent"], "labels": ["Kinase"]}],
"edges": [
{"type": "parent", "source": 'n_8', "target": 'n_7'},
{"type": "parent", "source": 'n_10', "target": 'n_9'},
{"type": "parent", "source": 'n_11', "target": 'n_9'},
{"type": "link", "source": 'n_12', "target": 'n_10'},
{"type": "link", "source": 'n_11', "target": 'n_8'}],
"main_nodes": ['n_9']}
], {
"comment": "B protein mutation description",
"nodes": [
{"id": 'n_13', "types": ["component", "agent"], "labels": ["B"]},
{"id": 'n_14', "types": ["component", "flag"], "labels": ["phos"], "values": ['u', 'p']},
{"id": 'n_15', "types": ["component", "region"], "labels": ["reg1"]},
{"id": 'n_16', "types": ["component", "keyres"]},
{"id": 'n_17', "types": ["attribute"], "labels": ["#aa"], "values": ['t', 'y']},
{"id": 'n_19', "types": ["attribute"], "labels": ["#loc"], "values": [17, 120]},
{"id": 'n_18', "types": ["attribute"], "labels": ["#loc"], "values": [72]}],
"edges": [
{"type": "parent", "source": 'n_15', "target": 'n_13'},
{"type": "parent", "source": 'n_16', "target": 'n_15'},
{"type": "parent", "source": 'n_19', "target": 'n_15'},
{"type": "parent", "source": 'n_17', "target": 'n_16'},
{"type": "parent", "source": 'n_18', "target": 'n_16'},
{"type": "parent", "source": 'n_14', "target": 'n_16'}],
"main_nodes": []}
}]
}

```