Appendix 2: Formal Modeling and Illustrative Example of a QoR-based Resource Graph

The functions required for user request define with their dependencies, a Workflow Model, WM, with WM $\subset$ FG. Based on the functions order in WM, the resources identified during the discovery process are linked together, forming a Directed Resource Acyclic Graph, DRAG. Formally, DRAG is defined as:

Definition 1 $\boldsymbol{DRAG = (DRES, Rel, f_{DRES}, f_{Rel})}$*, where:*

- $\boldsymbol{DRES}$*, is the set of the discovered static and dynamic resources obtained from the resource discovery process.*
- $\boldsymbol{Rel}$*, is the set of relations linking the resources together.*
- $\boldsymbol{f_{DRES}}$*, is the function computing the score of each resource function based on QoR values (e.g., Availability and Cost), included in our study in Hydra-based resources description.*
- $\boldsymbol{f_{Rel}}$*, is the function linking the resources together based on WM, and computing their link score based on their I/O similarities.*

The resources discovered for the same function, form a resource group, RG_f, relative to that function, where: $\boldsymbol{RG_f} = \bigcup_{\boldsymbol{i=1}}^{\boldsymbol{m}} \{\boldsymbol{res_{(f,i)}}\}$, with m is the number of candidate resources realizing function f, and $\boldsymbol{res_{(f,i)}}$ refers to the resource $\boldsymbol{res_i}$ providing f. A resource composition, RC, consists of a set of resources included, each, in a different $\boldsymbol{RG_f}$, where: $\mathbf{RC} = \bigcup_{\boldsymbol{f=1}}^{\boldsymbol{n}} \{\boldsymbol{res_{(f,i)}}\}$, such that n is the number of functions in WM, and $\boldsymbol{i} \in \boldsymbol{m}$, with m denotes the number of resources in the correspondent $\boldsymbol{RG_f}$. During selection, I/O matching between linked eligible resources is computed, forming the score link of these resources. Such score is calculated as: $\text{sim}(\boldsymbol{res_{(f,i)}}, \boldsymbol{res_{(f',j)}}) = \sum_{\boldsymbol{u=1}}^{\boldsymbol{U}} \sum_{\boldsymbol{v=1}}^{\boldsymbol{V}} \boldsymbol{sim}(\boldsymbol{out_u^{res(f,i)}}, \boldsymbol{in_v^{res(f',j)}})$, with:

- $\boldsymbol{res_{(f,i)}}$, $\boldsymbol{res_{(f',j)}}$, denote resources that belong, respectively, to $\boldsymbol{RG_f}$ and $\boldsymbol{RG_{f'}}$, where f precedes f' in WM
- $\boldsymbol{out_u^{res(f,i)}}$, is an output of $\boldsymbol{res_{(f,i)}}$, and U is the number of $\boldsymbol{res_{(f,i)}}$ outputs
- $\boldsymbol{in_v^{res(f',j)}}$, is an input of $\boldsymbol{res_{(f',j)}}$, and V is the number of $\boldsymbol{res_{(f',j)}}$ inputs

In our work, the matching score between an output of a res and an input of another, can be computed using any similarity measure function between keywords (as Jaccard measure [1]), and such that sim($\boldsymbol{res_{(f,i)}}$, $\boldsymbol{res_{(f',j)}}$) $\in$ [0, 1]. Each RC in DRAG has a score, score(RC), computed using (i) the score of each involved resource providing the needed function, score($\boldsymbol{res_f}$), and (ii) the score of the link relating each 2 eligible resources, such that: **score(RC) = Score(RES) + Score(Rel)**, where:

- **Score(RES)** = $\sum_{f=1}^{n} \boldsymbol{score(res_{(f,i)})}$, is the sum of the scores of the involved resources realizing the required functions, such that: n is the total number of functions in WM and i $\in$ m, with m denotes the number of candidate resources in the correspondent $\boldsymbol{RG_f}$.
- **Score(Rel)** = $\sum \boldsymbol{sim(res_{(f,i)}, res_{(f',j)})}$, is the sum of I/O similarity scores of each 2 eligible linked resources in RC, where: f precedes f' in WM, i $\in$ [1,m] and j$\in$ [1,m'], with m and m' denoting the numbers of resources in $\boldsymbol{RG_f}$ and $\boldsymbol{RG_{f'}}$ respectively.

Score(RES) and Score(Rel) can be multiplied, each, by a weight value defined in W in user request, r, allowing users to assign them a priority during compositions score calculation.

An example of a DRAG formed by discovered resources is given in Figure 1. As per the illustrated Figure, and based on the Workflow model showing the dependencies between the required functions to realize $\boldsymbol{f_5}$, the identified resources during resource discovery process are grouped into the same group relative to their provided same function. Each resource has a score computed according to the set of the quality attributes values related to its provided necessary function, i.e., $\boldsymbol{vf_i}$, and the attributes related to the resource itself, i.e., $\boldsymbol{vres_i}$. Also, each link relating two resources (e.g., $\boldsymbol{res_{3,2}}$ and $\boldsymbol{res_{5,2}}$) has a similarity measure score (e.g., $\boldsymbol{sim(res_{(3,2)}, res_{(5,2)})}$ between $\boldsymbol{res_{3,2}}$ and $\boldsymbol{res_{5,2}}$). A score(RC) is assigned to each possible resource composition in DRAG, which is represented by a path (see the red circled resources in Figure 1) linking one resource belonging, each, to different resource group.

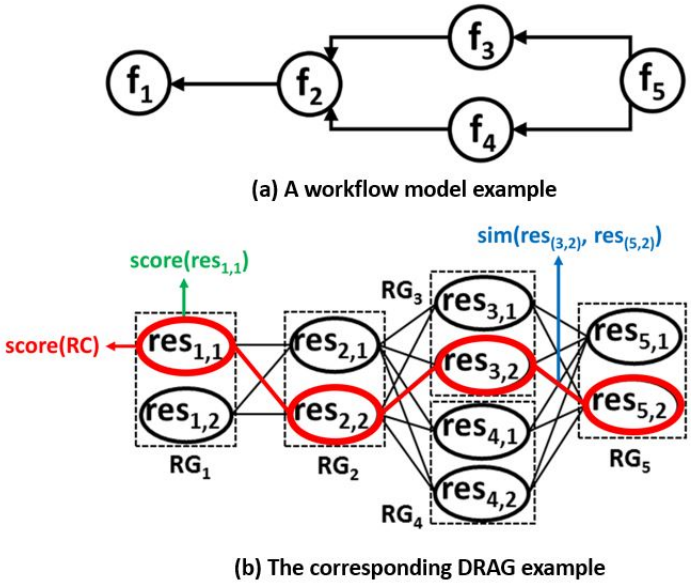

Figure 1: An example of a DRAG showing the scores defined for each of the involved resources, their I/O matching, and each possible composition

References

[1] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu. Using of jaccard coefficient for keywords similarity. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 380–384, 2013.