APPENDIX A

Already, distinct works have been done for conducting SMS and designing a unique research methodology [5, 10]. One of the most prominent of these research methodologies is related to Peterson et al. [5]. The research methodology used in this paper is adapted from three SMS that were done by Peterson et al. [5], Ramaki et al. [11], and Javan et al. [6] and some updates and improvements in some phases have been done. Generally, the suggested SMS consists of two main steps i.e. planning step and conducting step. Each of these steps is completely explained in the rest of this section.

1 PLANNING THE MAPPING STUDY

As can be seen in Fig.A. 1, the presented SMS consists of different steps. At the first step, the scope and research questions are specified and then three steps, namely planning the mapping study, evaluating the mapping study, and conducting the mapping study steps are simultaneously executed. Each step has a specified strategy to perform the mapping study that in the following, has been described in detail.

1.1 Specifying the Scope and Research Questions (RQs)

Defining some research questions in a review research work is essential. These research questions should be responded during the review by the researchers. Indeed, the objective of defining RQs is to determine the research goals. Therefore, one of the most important steps in an SMS process is designing some useful questions so that after answering them, all challenges, issues and important topics in the desired field be determined. In this SMS, 8 comprehensive and different RQs have been defined, which responding them can cover all our objectives. Table 1 in article describes these RQs and explains the rationality of each of them.

1.2 Specifying the Search Strategy

To conduct the search process, specifying the exact search strategy is vital which should be carried out during the process of review. In this paper, as previously stated, three distinct search strategies (i.e. database search, backward snowballing search, and manual search) have been designed to find all related studies. It's worth mentioning, considering the investigation of research studies, the advent of the cloud broker is 2009. Therefore, this SMS covers all published research articles from 2009 until end of 2019 and a complete review have been conducted on the cloud broker research works during this period. At the first step, all related papers from 2009 to 2019 are retrieved by combining the results of backward snowballing and manual searches. Moreover, a distinct search on well-known databases like Google Scholar, Springer Link, IEEEXplore, ACM Library, and ScienceDirect is done to find unseen related articles and to identify the possible shortcoming in the snowballing and manual search strategies.

Fig.A. 2 illustrates the general strategy which comprises seven steps. It should be noted that the determining related search spaces (journals, conferences, and workshops) are not necessarily carried out after the manual search step. Hence, an S-Flag is defined and set with zero in step 2 and one in step 5. Current search strategy comprises seven primary steps that are acquired by investigating the base reviews studies [5, 6, 11] and applying some improvements to them.

- In this step, a set of initial secondary studies (see Table 1 in article) are randomly selected and investigated to determine a set of initial keywords that are used in the next steps for finding the related studies. Also, a phase variable is used to show the number of passed steps.
- In this step, for finding the related studies with high quality, two types of quality measures are specified that are called qualification and relation criteria. As mentioned previously, the search strategy should be able to extract and find all related search spaces (journals, conferences, and workshops) to retrieve all studies related to the broker topic. Therefore, to arrive at the high-quality studies, the existence of quality

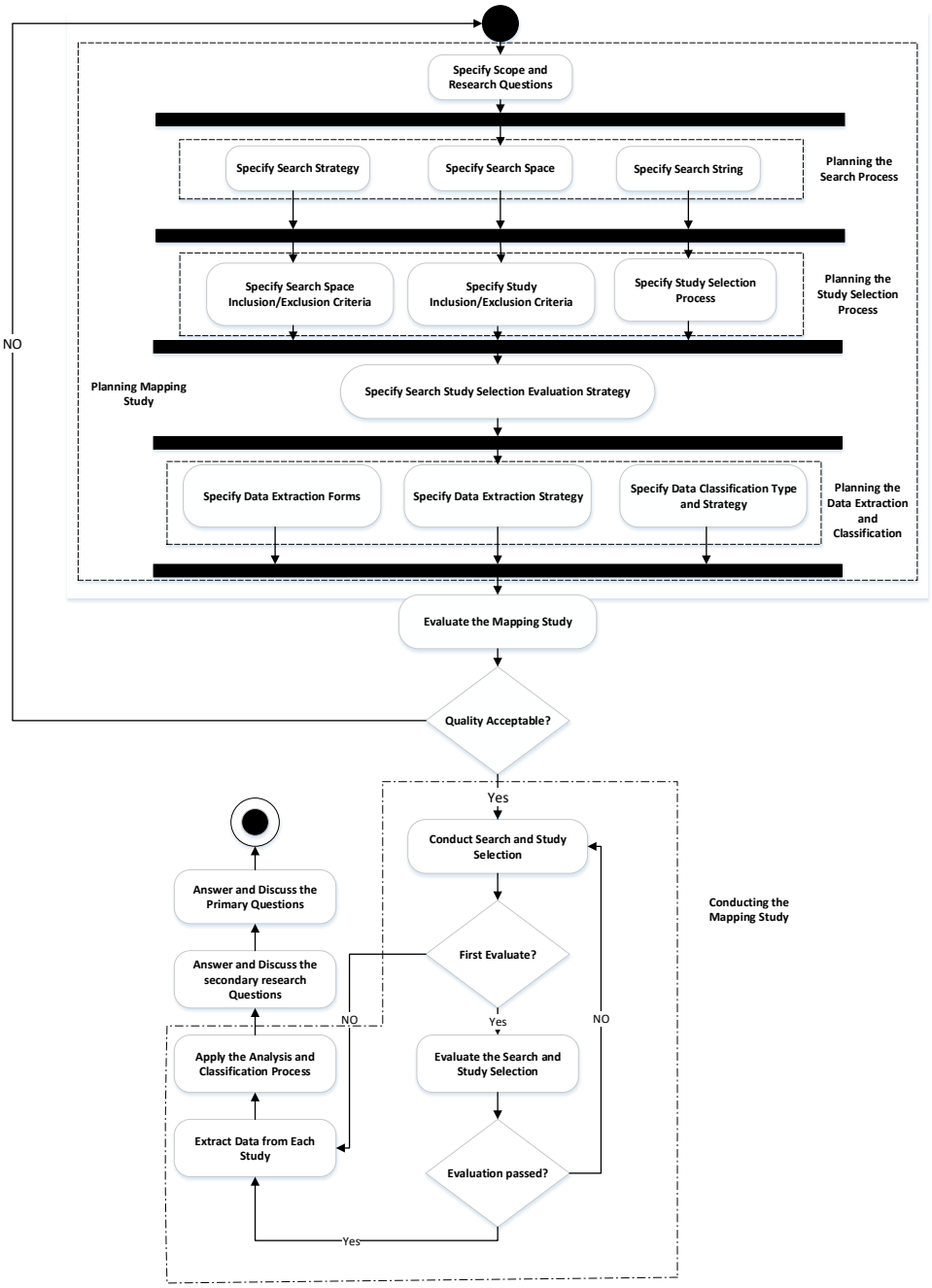

Fig.A. 1. Research methodology [6]

measures is necessary that filters journals, conferences, and workshops namely, search space, in terms of quality factors defined in the SMS. To reach a set of initial search spaces, all cited studies in the initial secondary studies are extracted. After reviewing all initial secondary studies, if a new search space is found, then step 3 should be done, else the search process will terminate and step 5 should be run to conduct the backward snowballing on the included studies. Consequently, a flag called S-Flag has been defined to prevent ineffective searches during the search process. If the search process is in the backward snowballing step, the amount of S-Flag is equal to one, else it is set to zero. S-Flag in this step will be initialized with zero. When a backward snowballing is done on the included papers, the S-Flag will be set to one.

- In this step, the extracted search spaces should be evaluated considering two defined quality measures and the compatible search spaces should be selected. The set of search spaces is updated by adding the extracted new search spaces. The output of this step is a set of high-quality related search spaces that concludes all related papers in the field of cloud broker. The manual search is done on the selected search spaces by the initial set of keywords corresponding queries defined in Table 2 in article to find the related papers.
- In this step, the study selection process is done. If a study satisfies all quality criteria, therefore, it is an included study and the search process jumps to step 5. Moreover, if there are some new keywords in the included studies, then they are extracted and added to the keywords set. This action concludes a complete set of related keywords in the field of cloud broker that is led to a comprehensive search study set. If the paper has been acquired of the backward snowballing, the process jumps to step 2.
- In this step, backward snowballing is done on the all new found included papers and S-Flag is set to one. If the backward snowballing results new papers therefore, the execution jumps to step 4 to conduct the search space process, otherwise step 6 is done.
- with applying a set of pre-defined keywords, a database search is done and new unseen papers are extracted.
- at the last step, the study selection process should be executed on the recently found papers and after that, the search process ends.

As previously mentioned, an auxiliary document called *SuppFile* has been provided and throughout the paper, may be referred to it. For conducting the database search, different databases for finding the related papers in the desired field have been investigated such as ScienceDirect, ACM Digital Library, IEEE Xplore, Springer Link, and so on. A complete list of extracted studies can be seen in $SuppFile_{E10,T1}$ to $SuppFile_{E10,T6}$ (Supplementary document in *SuppFile* folder). As previously mentioned in the description of the search strategy, some quality criteria for evaluation of search spaces quality and found papers have been applied. Therefore, the database search is done just on the high-quality search spaces that have been more related in terms of scopes. The $SuppFile_{W3,T4}$ to $SuppFile_{W3,T7}$ shows a complete list of all extracted search spaces in our research field.

1.3 Specifying the Search Space

At the beginning of conducting the SMS process, the search space set is empty. Therefore, after selecting some secondary studies some journals, conferences and workshops are added to the search space list by investigating articles cited in the references section of secondary studies.

1.4 Specifying the Search String

The objective of this SMS is to review all papers in the field of cloud broker. To search related papers in the manual and database search phases, a set of keywords have been defined that can be seen in $SuppFile_{W3,T3}$. These keywords are gradually completed during the progress of the review process. The keyword set is completed by the three mentioned phases. Also, as previously mentioned, some queries have been defined to search related studies in the database of different publishers. These queries are shown in Table 2 in article and are a combination of elementary defined keywords in the first step of finding papers.

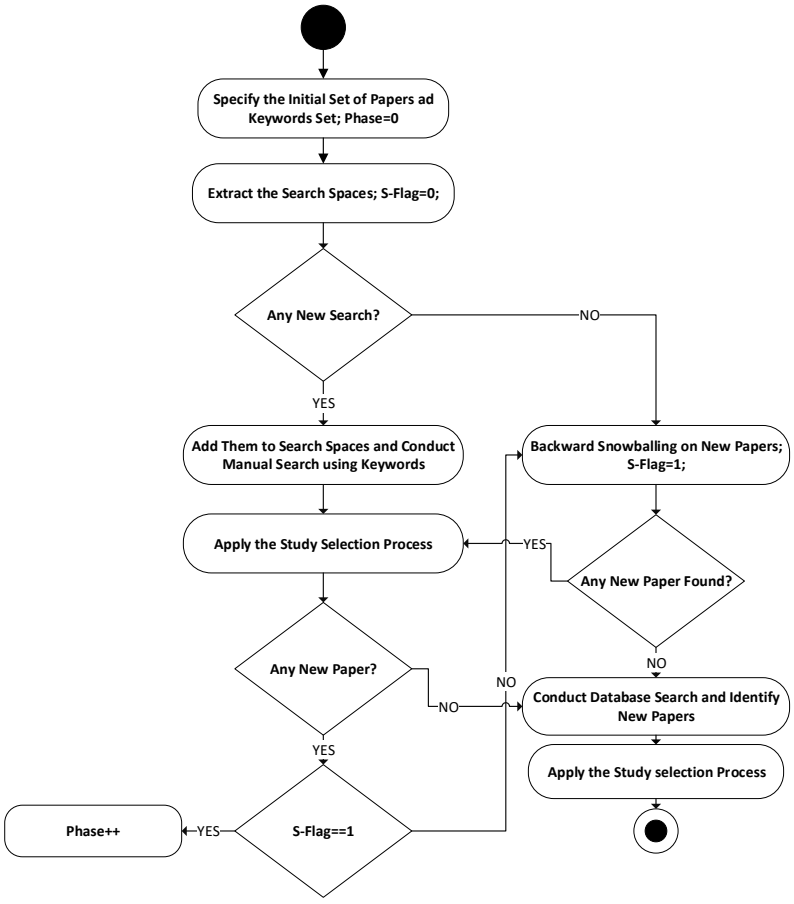

Fig.A. 2. The Search Strategy [6]

1.5 Planning the Study Selection Process

As shown in Fig.A. 1, this phase comprises three primary steps namely, *specifying the search space inclusion/exclusion criteria*, *specifying study inclusion/exclusion criteria*, and *specifying the study selection process*. In the following, each of the phases will be explained in detail.

•Specifying the search space inclusion/exclusion criteria.

For beginning the review process, a set of review and survey papers in the desired field are found and some of them are selected by experts. These selected review and survey papers that have been placed in Table 1 in article are called initial secondary studies. In the first step, all search spaces in the reference list of the secondary studies are investigated and extracted into a list of new search spaces. Then some search spaces with good quality and related scope are selected among them. To arrive at this goal, some quality criteria for determining the inclusion/exclusion state of search spaces have been specified. The exclusion criteria are illustrated in Table.A. 1

for journals and Table.A. 2 for conferences and workshops. The result of applying the exclusion criteria is removing all unrelated or low-quality search spaces.

Table.A. 1. Journal Exclusion Criteria (JEC)

# JEC	Description
1	If the journal is not indexed in the JCR
2	If the scope of the journal is not related to our desired field

Table.A. 2. Other Exclusion Criteria (OEC)

# OEC	Description
1	((Qualis<A5) OR (ERA<A) OR [(H5_Index<15) AND ((Qualis<A5) OR (ERA<A))] OR (Metrics Not Available))
2	(Aims and scopes are not related)

To evaluate the quality of journals, the JCR metric has been used and ERA and H5-Index metrics are used for conferences and workshops (these are shown in Table.A. 1 and Table.A. 2). It should be noted that there are some valuable points that are explained below:

- The field of cloud broker comprises a large pool of research studies, therefore some thresholds for the exclusion criteria considering two principals have been empirically selected: 1) a small change in the exclusion thresholds should not have a big effect on the number of excluded/included papers and 2) applying these thresholds should not exclude a lot of highly cited papers.
- Although the desired field in this SMS is cloud broker, but search spaces with the scope and aim of web services or distributed and parallel computing have been considered. There are also search spaces in other research areas which are included, because they have published some papers in the field of cloud computing. An example is IEEE Transaction on Smart Grid.
- There are some search spaces such as the International Conference on Software Engineering (ICSE) that have scarcely published papers in the field of cloud computing. Because its scope is about software engineering. Therefore, such search spaces are excluded.
- In rare cases, we found a search space with two names. For example, the ACM International Symposium on High-Performance Distributed Computing and International Symposium on High-Performance Parallel and Distributed Computing are two names for the search space of HPDC. Also, there is a search space that its name had been changed. The initial name for GLOBECOM search spaces is IEEE Global Telecommunications Conference, and after 1972, its name has been changed to Global Communications Conference. As another example, after 2007, two search spaces namely, European Conference on Machine Learning (ECML) and European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD) have been merged titled European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD).

•Specifying the study inclusion/exclusion criteria

After finding the related and included search spaces, they are searched using constructed queries in Table 2 in article to find the related papers. $SuppFile_{E1,T5}$ presents complete information of all included papers. The inclusion criterion for an extracted study is its relevance to our designated scope. To determine the inclusion/exclusion state of each extracted paper, title, abstract, and keywords have been generally investigated. In some cases, the entire paper has been reviewed to ensure its inclusion or exclusion status. A complete list of all excluded papers

can be seen in $SuppFile_{E2,T2}$. It is worth mentioning that all extracted review studies during conducting the search process in this SMS have been excluded. A complete list of all extracted review studies in the desired field can be seen in $SuppFile_{E2,T1}$. Table.A. 3 illustrates the exclusion criteria for extracted studies.

Table.A. 3. Exclusion criteria for extracted studies

#	Description
1	The study is not a primary study (e.g. survey)
2	Study cannot be accessed (e.g. not indexed)
3	The study is out of our primary scope(e.g. security)
4	The study belongs to an excluded search space
5	The contribution of the study is not related to the cloud broker.

By applying the exclusion criteria for search spaces and determining the included search spaces, the manual search for finding related studies are just conducted for the included search spaces. However, by using the backward snowballing search and database search some studies may be retrieved from the excluded search spaces. Therefore, for checking this situation, an exclusion criterion is defined (see Table.A. 3, row 4). According to the first criterion, secondary studies (for example reviews, surveys) have not been investigated in this SMS.

•Specifying the Study Selection Process

Fig.A 3 illustrates the study selection strategy that comprises two main steps. In the first step, the extracted studies are reviewed by reading their titles, abstracts, keywords and in some cases full-text. If the paper is irrelevant or its relevancy is uncertain, therefore the judgment of a third-party is needed to determine its relevancy. In the second step of the study selection strategy, determining the exclusion of paper is done by using the defined criteria in Table.A. 3.

1.6 Specifying the Search and Study Selection Evaluation Strategy

In this review, two prominent and most valid metrics are used to evaluate the search strategy. In other words, both objective evaluations (i.e. quantitative criteria) and subjective evaluation (managed by expert review) are done using Eq.(A.1)) and Eq.(A.2) respectively. To have a more objective evaluation, the quasi-sensitivity metric is used to evaluate the applied search and study selection.

$$Sensitivity = \frac{\text{The Number of Studies in Our SMS}}{\text{The Number of Studies Overall}} \times 100 \quad \text{(Eq.(A.1))}$$

$$QGS = \frac{\text{number of discovered papers in the search phase}}{\text{number of discovered papres in the evaluation phase}} \times 100 \quad \text{(Eq.(A.2))}$$

QGS alludes to a set of studies published in the well-known research communities. To create this gold standard, the home pages of active researchers in the area of cloud computing have been visited and their papers in the cloud broker field have been extracted. After extracting studies from home pages and applying the inclusion/exclusion criteria, the remaining unseen papers comprise our GCS according to Eq.(A.2). The aim of applying the QGS is calculating the quasi-sensitivity and then comparing the obtained result with a predefined threshold. Accordingly, if the result falls below the threshold, the search and study selection process should be repeated using QGS. By following [2], an acceptable threshold should be between 70% and 80%. In the evaluation phase, 32 articles were found, of which 26 articles were found in the main search phases of the systematic review process, therefore, we achieved a sensitivity of 81.25% which is above our predefined threshold. It means that the probability of not

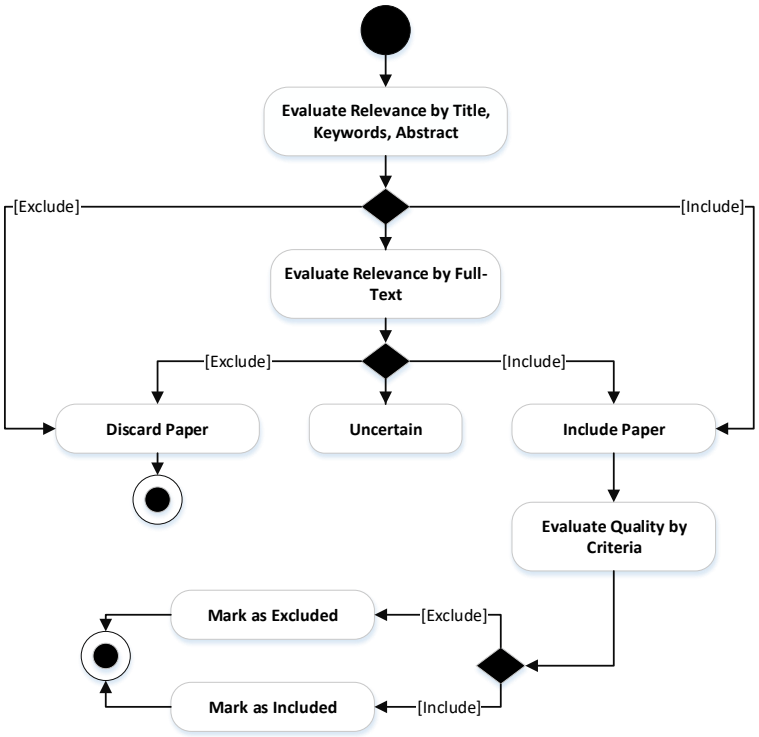

Fig.A. 3. The study selection strategy [6]

finding a paper related to the cloud broker is less than 20%. Therefore, it can be concluded that the acquired results from this review have enough accuracy and validity.

1.7 Planning the Data Extraction and Classification Process

After conducting the study search process and determining the included papers, to answer the RQs, some information is needed that should be extracted from these papers. Table 8 in article illustrates that which information is useful for answering which RQs. It should be noted that the "Topic" word in Table 8 in article represents an acquired concept from the "keywords" field using a keywords clustering algorithm that is a common approach for topic-dependent classification [5]. $SuppFile_{E(1,5to9)}$ include comprehensively all needed information to answering RQs.

2 EVALUATING THE MAPPING STUDY

The evaluation criteria that have applied to evaluate this SMS are similar to [5]. Details of this information can be seen in $SuppFile_{W3,(T4toT7)}$. During the review process, if the quality level is inadequate, the search process must be revised to solve existing problems.

Table.A. 4. Sample Extracted Journal Search Spaces and Results of the Search Space Selection Process

#JID	Journal Name	ISSN	$IF_{JCR-2017}$	$Q_{JCR-2017}$	Topic	Publisher	Depth	Reason of Exclusion
6	Future Generation Computer Systems	1872-7115	4.639	Q1	Computer science a Theory and methods	Elsevier	0	-
7	IEEE Internet Computing	1941-0131	1.929	Q1	Computer science a Software engineering	IEEE	0	-
11	Annals of Telecommunications	1958-9395	1.168	Q4	Telecommunication	Springer	0	JEC2
12	IEEE Internet Computing	N/A	-	-	N/A	Springer	0	JEC1

3 CONDUCTING THE MAPPING STUDY

As can be seen, the applied strategies for this SMS have been described in Sections 1 and 2 In the following, the obtained results of applying these strategies have been shown.

3.1 Conducting the process of Search space and study Selection

The first step when running the search process is providing an initial set of papers that are acquired from a set of secondary studies that have been found using an informal search process by using the keywords *cloud broker* along with keywords such as *survey* and *review*. A *depth* variable describes the type of search in our search methodology. After acquiring a set of secondary studies (shown in Table 1 in article), the initial value of depth is set to zero and related cited papers from those studies are extracted that constitute an initial set of included studies (primary studies). After that, the backward snowballing and manual search is done on the included papers in the initial set. After finishing manual and snowballing search, for complementing the results set, a database search is done. Table.A. 4 and Table.A. 5 demonstrate sample records of journals, conferences and workshops respectively that their inclusion/exclusion state have been determined based on the exclusion criteria defined in Table.A. 1 and Table.A. 2. The *Reason of Exclusion* column of these tables determines the reason for exclusion for each search space.

For example, in Table.A. 4, the journal with JID = 11 is excluded since its scope is not related to the defined scope of the cloud broker. As another example, the journal with JID =12 is excluded, because it has not been indexed in JCR. Table.A. 5 illustrates that a conference with OID = 52 is excluded because its metrics i.e. Qualis and ERA are less than a defined amount in Table.A. 2. Also, OID=62 is excluded because its scope is not related to the scope of the cloud broker field (refer to $SuppFile_{W3,T1}$). Complete information about all extracted search spaces and reason for exclusion and inclusion of journals, and other (conferences, and workshops) have been provided in $SuppFile_{W3,T5}$ and $SuppFile_{W3,T7}$ respectively. Table.A. 6 and Table.A. 7 demonstrate a sample of extracted papers based on the described search spaces in Table.A. 4 and the exclusion and inclusion states for each paper considering the exclusion criteria has been determined. For example, in Table.A. 6, the presented contribution by the first study is not related to the aim and scope of a cloud broker, so this study is marked as an excluded paper.

$SuppFile_{W3,T4}$ and $SuppFile_{W3,T6}$ present list of extracted papers and comprehensive information about the reason for the exclusion of each study. The main aims and scope of the cloud broker field in $SuppFile_{W3,T1}$

Table.A. 5. Sample Extracted other Search Spaces and Results of the Search Space Selection Process

#OID	Other[1] Name	H5_Index	Qualis	ERA	Indexing	Depth	Reaseon of Exclusion
46	ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)	77	A1	A	ACM	0	-
52	International Conference on High Performance Computing and Communications (HPCC)	20	-	B	IEEE	0	OEC1
62	International Conference on Information and Knowledge Management (CIKM)	49	-	A	ACM	0	OEC2
70	International Conference on Data Engineering (ICDE)	53	A1	A	IEEE	0	-

[1]Conference, Workshop and Symposium

Table.A. 6. Sample extraction table for journal papers

Paper Title	Journal Name	Year	Exclusion Criteria
Privacy-preserving and sparsity-aware location-based prediction method for collaborative recommender systems	Future Generation Computer Systems	2019	PEC1
A trust centric optimal service ranking approach for cloud service selection	Future Generation Computer Systems	2018	-
A privacy-preserving cryptosystem for IoT E-healthcare	Information Sciences	2019	PEC1
Implementation of a real-time network traffic monitoring service with network functions virtualization	Future Generation Computer Systems	2019	PEC1
A hybrid multi criteria decision method for cloud service selection from Smart data	Future Generation Computer Systems	2019	-

have been determined. Moreover, the aims and scope of each of the search spaces have been introduced in $SuppFile_{W3,T5}$ and $SuppFile_{W3,T7}$ and the reason for the exclusion of each search space has been determined in them.

After constructing Table.A. 6 and Table.A. 7 and determining the included studies, by investigating the keywords of the included papers, some new keywords are extracted and added to the set of keywords for later use. $SuppFile_{E10,T4}$ and $SuppFile_{E10,(T5toT6)}$ illustrate complete information of all other papers (conferences and workshops) and journal papers respectively. $SuppFile_{E1,T5}$ presents a complete list of all included papers (journals, conferences, and workshops).

3.2 Evaluating the Search and Study Selection Process

The obtained results of the evaluation phase ($SuppFile_{E1,(T1toT5))}$ have been used to extract the active researchers. A list of active researchers can be seen in $SuppFile_{E1,T2}$ and the author's country can be found in $SuppFile_{E1,T1}$.

Table.A. 7. Sample extraction table for Conference papers

Paper Title	Conference Name	Year	Exclusion Criteria
A model for evaluating the economics of cloud federation	International Conference on Cloud Networking (CloudNet)	2015	-
Cloudlet Scheduling with Particle Swarm Optimization	International Conference on Communication Systems and Network Technologies (CSNT)	2015	PEC1
Incentivizing Microservices for Online Resource Sharing in Edge Clouds	International Conference on Distributed Computing Systems (ICDCS)	2019	-
The Elasticity and Plasticity in Semi-Containerized Co-locating Cloud Workload: A View from Alibaba Trace	ACM Symposium on Cloud Computing	2019	PEC1

The validation of this SMS has been evaluated by conducting the evaluation phase. In this phase, the homepage of authors manually have been probed for finding unseen papers during the execution of three search phases. $SuppFile_{E10,T7}$ illustrates a list of unseen papers during conducting search phases. Based on the provided information, there are 6 included studies that our SMS failed to find them. In other words, during the review process, their search space has not been taken into consideration. Again, the search strategy was done on these unexplored studies that led to some new included search spaces (4 journals, 18 conferences and 7 workshops) to the search spaces set. The investigation of these new search spaces added 17 new included studies to the studies set. Considering these conducted processes to evaluate the completeness of the search strategy, we conclude that the set of studies has satisfactorily covered the cloud broker field.

3.3 Data Extraction

After completing the set of studies, the process of data extraction and classification is performed on them.

3.4 Analysis and Classification of data

After organizing data into tables, to respond to the RQs of our SMS, these data should accurately be investigated and analyzed. The most important objective in the presented SMS is determining the primary topics and sub-topics in the cloud broker field. In the rest of this section, we explain the process of topic and sub-topic detection.

So far, some techniques have been presented to identify topics in the research fields. Detail information about this phase is placed in $SuppFile_{W1,T1andT2)}$.Among all proposed approaches, an outstanding approach that has been proposed by [7, 12] applies a statistical similarity factor to classify keywords and discovers the characterizations of topics. The main disadvantage in this approach is the determination of topics and sub-topics just according to a small piece of the paper. For example, they just consider a paragraph of the paper to determine the topic. In this review to determine the topic of a paper, we usually investigate the whole paper.

Usually, the similarity-based keywords clustering techniques do not have a high classification power [4], therefore, other techniques are proposed by some researchers [1, 3, 13] that can group the keyword set into some clusters based on co-occurrences matrix of keywords, and find the related topics and sub-topics considering their similar cognitive orientation [13]. Besides, in each step, the number of clusters corresponds to the determined issues in the SMS. In these approaches, detection of topics is heuristically done, but in this SMS, we propose a new approach based on co-occurrences matrix and use the knowledge of expert/experts for finding topics and subtopics. Previously, we have used this technique in our review works [4, 9].

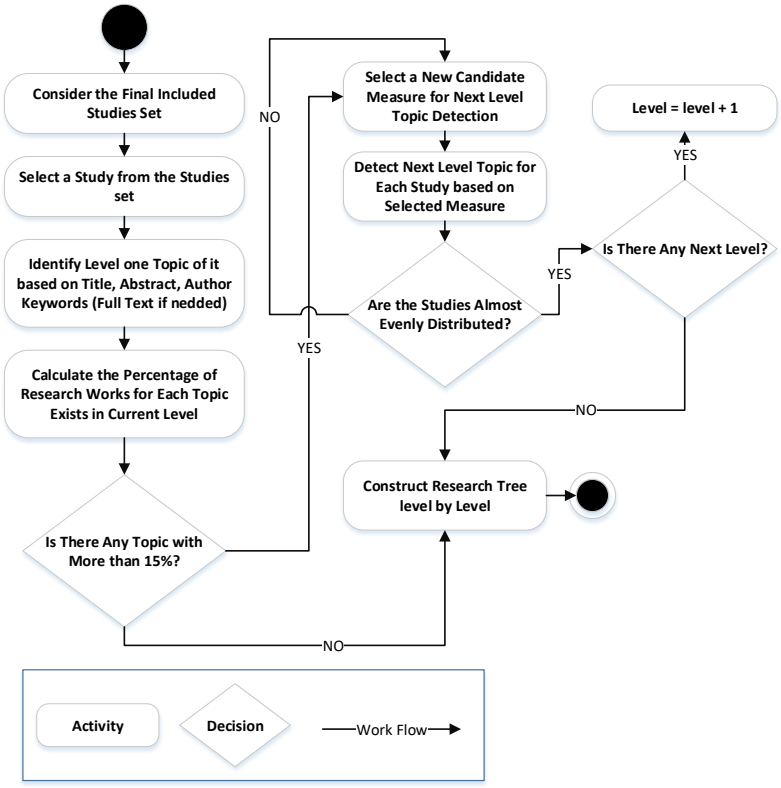

Fig.A. 4. The construction process of the research tree [11]

The last step in the search process is to identify a set of keywords from included papers that may be done directly or indirectly. When keywords set are available in an included paper, these keywords are chosen, otherwise, the extraction of keywords must be done manually by an expert. Fig.A. 4 demonstrates the construction process of the research tree. A research tree is a multilevel tree that contains the main topics and sub-topics in the desired field. The most important level in this tree is its root. Because all analysis to respond to RQs is done based on the located topics and sub-topics into the root level. The extraction process of level-one topics in research tree can be seen in Fig.A. 5.

In this figure, first all keywords of included papers should be extracted and then the most frequent keywords which are appeared more than a threshold value are chosen. Then, the co-occurrence matrix should be constructed from the frequent keywords. A sample of the co-occurrence matrix is shown in Table.A. 8. xij is a cell in the ith row and the jth column and illustrates the number of times that the word of the ith row appears with the jth column word at the same paper. For example, the service composition keyword appears in 88 studies, which co-occurs with QoS keyword in 31 studies. A complete co-occurrence matrix that comprises 40 keywords is placed in $SuppFile_{E3,T1}$. After Construction of co-occurrence matrix to detect the cognitive orientation of two keywords in relation to all other keywords, it should be normalized and the similarity of each keyword pair on the co-occurrence matrix should be computed. In other words, each keyword has a vector that each element of

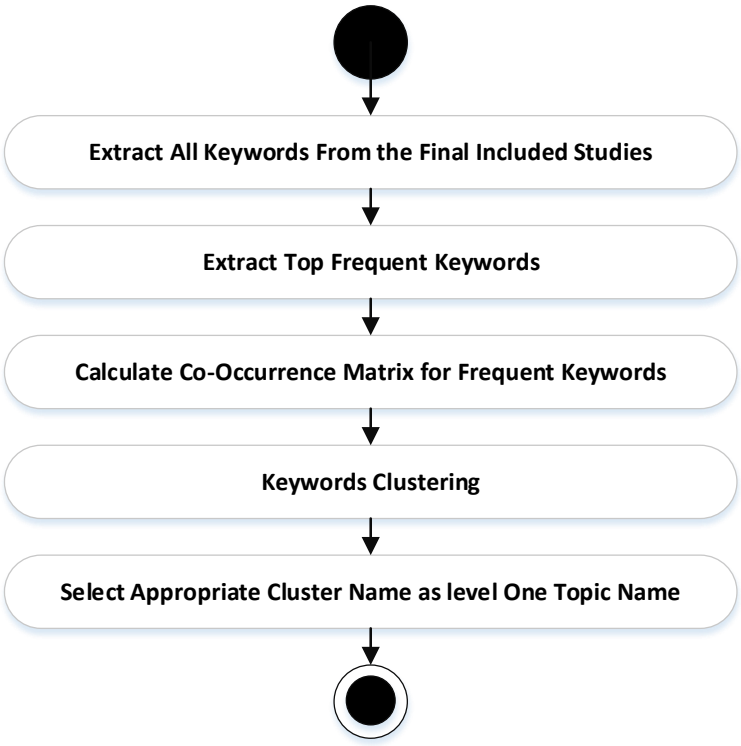

Fig.A. 5. The extraction process of level-one topics in research tree [11]

Table.A. 8. A sample of the co-occurrence matrix

Keywords	Service Composition	QoS	Service Selection	Resource management	Scheduling
Service Composition	88	31	12	3	0
QoS	31	68	11	7	4
Service Selection	12	11	56	1	1
Resource management	3	7	1	51	10
Scheduling	0	4	1	10	32

it depicts co-occurrences with another keyword. The similarity is computed based on the cosine index, which is demonstrated in Eq(A.3) [13]. The resulting matrix should be classified. We use the K-means algorithm [8] as a common clustering method where K is initialized by an expert and illustrates the number of clusters to be constructed.

$$Sim(x,y) = \frac{\sum_i (x_i y_j)}{\sqrt{((\sum_i x_i^2)(\sum_i y_i^2))}} \quad \text{(Eq(A.3))}$$

The number of included papers is 496 that can be seen in $SuppFile_{E1,T5}$. In the following, we explain the process of research tree construction and top-level topic detection considering frequency of keywords.

- the keywords of all the included papers (496 papers) are collected. If the paper does not have the author's keywords, the keywords are produced by an expert. In the keyword's column in the $SuppFile_{E1,T5}$, if a keyword has been manually extracted by an expert, the cell color is yellow. To determine keywords manually by an expert, the title, abstract, and context of the paper and also the defined queries in Table.A. **??** have been used. The keywords of included papers have been collected in a keyword pool and then the frequency of each keyword has been computed. For example, the broker keyword has been seen in 100 included papers. The size of the keyword pool after the first phase was 1291.
- we have categorized keywords that are lexically similar. For example, SaaS, SaaS applications, and software as a service are located in the SaaS group. After grouping all keywords, the number of keywords decreased to 735 items.
- Some of the most repetitive keywords that do not give rise to a new concept in the research tree have been removed from the keyword list. Example of these keywords are broker, cloud computing, edge computing, distributed system, IoT, mobile cloud computing, and autonomic computing.
- The co-occurrence matrix is obtained for words with a repetition threshold greater than or equal to 9. The repetition threshold is obtained by trial and error, and at threshold 9 the best semantics is coincidentally found in the matrix output. A total of 35 keywords with a repetition greater than or equal to 9 formed the primary matrix. Then, in each iteration, the two words that have the most coincidence compared to other pairs of words, are selected and form a new subject. This process continues until there is no new co-occurrence in the matrix members. After 18 repetitions of the process of finding the highest number of coincidences, the final number of 18 items has been obtained. Table.A. 10 shows the compounds obtained as a result of the co-occurrence matrix and the topics obtained as the level-one topics of the research tree. By examining the related topics by experts and aggregating them, finally 10 main topics (the second column in Table.A. 10) have been created in the cloud broker field that is introduced as the level-one topics of the research tree. Topics are divided into two sections: client-centric and provider-centric.

Client-Centric topics are activities that the broker performs as a result of the user request. For example, as a result of a service request by a user, the broker may perform service discovery, service selection, service composition, and so on. Provider-Centric topics are activities that the broker performs as a result of the provider's request. For example, pricing, resource allocation, energy management, etc. It should be noted that service allocation includes service provisioning and scheduling on the client-side. Figure 1 in article shows the research tree obtained in this SMS. The percentage of included papers in each topic is shown below the topic. It should be noted that in the process of determining the subject of papers, thematic similarities are considered as Table.A. 9. In Table.A. 9, in each row, each topic in the topic column is equivalent with words inserted in the similar concept column.

Table.A. 10 shows the compounds obtained as a result of the co-occurrence matrix and the topics obtained as the level-one topics of the research tree. Based on Section 2, one of the rubrics applied for evaluation of our SMS is threats to validity. In the validation process, the primary goal is providing some evidence for resolving all existing threats facing our SMS. In the following, the prominent evidence is investigated and discussed.

- **Obtaining a set of high-quality studies:** for obtaining a complete set of high-quality studies in the field of the cloud broker, a complete procedure is designed as a search strategy that comprises advantages of both famous search methods i.e. SLR and SMS. Therefore, we believe that our review is reliable.

Table.A. 9. Thematic similarities between the keywords of included papers

Topic	Similar Concept
Service composition and integration	Composition + Orchestration + Integration
Service allocation	Scheduling + Provisioning
Energy management	Energy management + Green computing
SLA management	SLA violation + SLA management and etc.
Resource allocation	Resource allocation + Resource management + VM scheduling
Pricing	Pricing + Cloud market
Recommendation	Recommendation + Data management

Table.A. 10. The topics obtained as the level-one topics of the research tree

NO	Topic	Integrated Keywords
1	Service Discovery	Discovery Cloud servicePrediction
2	Service Allocation	Scheduling/Provisioning
3	Energy Management	Virtualization/Green Computing
4	Service Selection	Service Selection/Graph-based Methods
5	SLA management	SLA/Trust
6	Resource Allocation	Resource Allocation/IaaS Resource Management/Data Center
7	Pricing	Pricing/Game Web Service/Cloud Market
8	Monitoring	Cloud Provider/Monitoring
9	Recommendation	Recommendation
10	Other	Particle Swarm Optimization/Cloud Manufacturing Middleware/SaaS IoT/Genetic Algorithm Distributed Systems/Learning
11	Service Composition and Integration	Ontology/Optimization/Service Composition/QoS Integration

- **Obtaining the most related studies:** one of the most primary advantages in our search strategy is gradual evolution of keywords set during conducting the SMS. In should be stated that in some situation, some keywords in the set of keywords did not convey a concept of the cloud broker. Accordingly, all such keywords in a category by applying some logical operators (AND, OR) have been merged. Examples of merging some keywords are provided in $SuppFile_{E3,T2}$.
- **Reviewer's biases or misunderstandings during the process of study selection:** To prevent these challenges in our SMS, at first, the selection process is independently done by two reviewers, then the possible disagreements were solved by the third reviewer and decision rules (see Section 1.5).
- **Creating some forms to extract raw-data:** as another threat, during the execution of the data extraction phase, some included studies were without author's keywords and we extracted suitable keywords for those studies considering the context of it and stored the extracted keywords in the related forms.

- **Assigning proper name for each level-one topic:** After clustering all keywords (extracted and generated), we assign a proper name for each cluster that describes the concept of that cluster. For example, for the *Resource Allocation* topic, there were some keywords like *Resource Allocation*, *Resource Management*, and *VM Scheduling*. However, we pick out *Resource Allocation* as a suitable name for the topic, because we believe this name can goodly convey the desired concept of studies. Nevertheless, in the presented SMS, in addition to both considering semantic and syntax of keywords for each topic, we also brought up the selected names between our team to remove the potential bias.

REFERENCES

[1] Douglas G Altman and J Martin Bland. 1994. Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal* 308, 6943 (1994), 1552.

[2] Fariba Aznoli and Nima Jafari Navimipour. 2017. Cloud services recommendation: Reviewing the recent advances and suggesting the future research directions. *Journal of Network and Computer Applications* 77 (2017), 73–86.

[3] Nicolò Maria Calcavecchia, Antonio Celesti, and Elisabetta Di Nitto. 2012. Understanding decentralized and dynamic brokerage in federated cloud environments. In *Achieving federated and self-manageable cloud infrastructures: theory and practice*. IGI Global, 36–56.

[4] Sameer Singh Chauhan, Emmanuel S Pilli, Ramesh Chandra Joshi, Girdhari Singh, and Mahesh Chandra Govil. 2019. Brokering in interconnected cloud computing environments: A survey. *J. Parallel and Distrib. Comput.* 133 (2019), 193–209.

[5] Frank Fowley, Claus Pahl, Pooyan Jamshidi, Daren Fang, and Xiaodong Liu. 2016. A classification and comparison framework for cloud service brokerage architectures. *IEEE Transactions on Cloud Computing* 6, 2 (2016), 358–371.

[6] Abbas Javan Jafari and Abbas Rasoolzadegan. 2020. Security patterns: A systematic mapping study. *Journal of Computer Languages* 56 (2020), 100938.

[7] Ruhani Jain, Tejinder Sharma, and Narinder Sharma. 2017. A review on service broker algorithm in cloud computing. *International Journal of Computer Applications* 975 (2017), 8887.

[8] Hang Li and Kenji Yamanishi. 2002. Text classification using ESC-based stochastic decision lists. *Information processing & management* 38, 3 (2002), 343–361.

[9] Hang Li and Kenji Yamanishi. 2003. Topic analysis using a finite mixture model. *Information processing & management* 39, 4 (2003), 521–541.

[10] Nagarajan Rajganesh and Thirunavukarasu Ramkumar. 2016. A review on broker based cloud service model. *Journal of computing and information technology* 24, 3 (2016), 283–292.

[11] Ali Ahmadian Ramaki, Abbas Rasoolzadegan, and Abbas Ghaemi Bafghi. 2018. A systematic mapping study on intrusion alert analysis in intrusion detection systems. *ACM Computing Surveys (CSUR)* 51, 3 (2018), 1–41.

[12] Adel Nadjaran Toosi, Rodrigo N Calheiros, and Rajkumar Buyya. 2014. Interconnected cloud computing environments: Challenges, taxonomy, and survey. *ACM Computing Surveys (CSUR)* 47, 1 (2014), 1–47.

[13] He Zhang, Muhammad Ali Babar, and Paolo Tell. 2011. Identifying relevant studies in software engineering. *Information and Software Technology* 53, 6 (2011), 625–637.