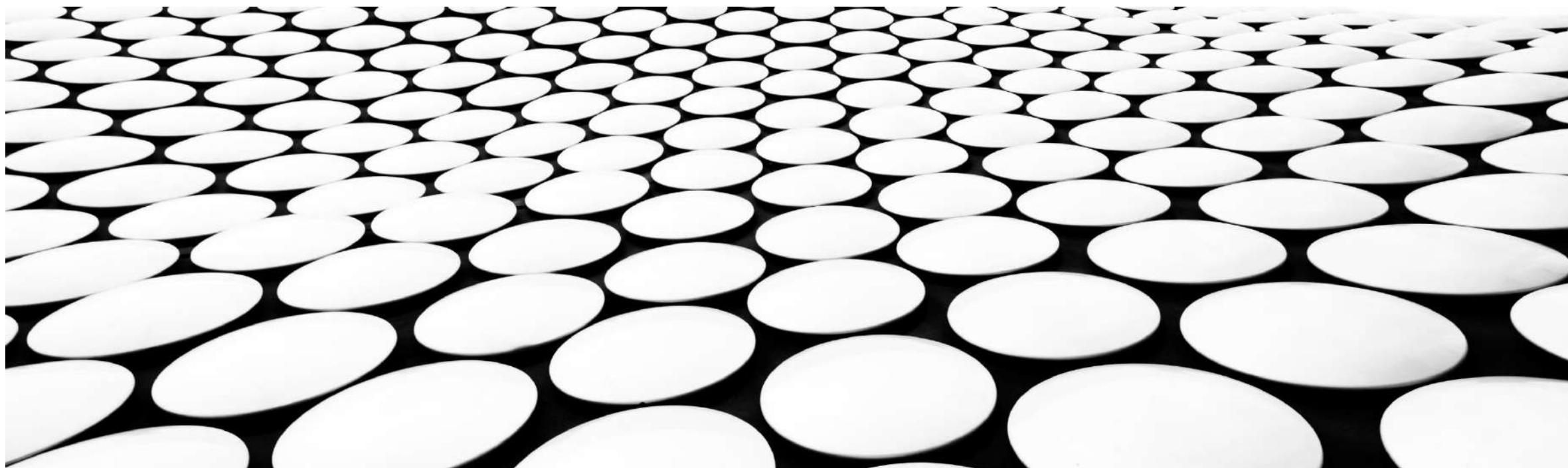


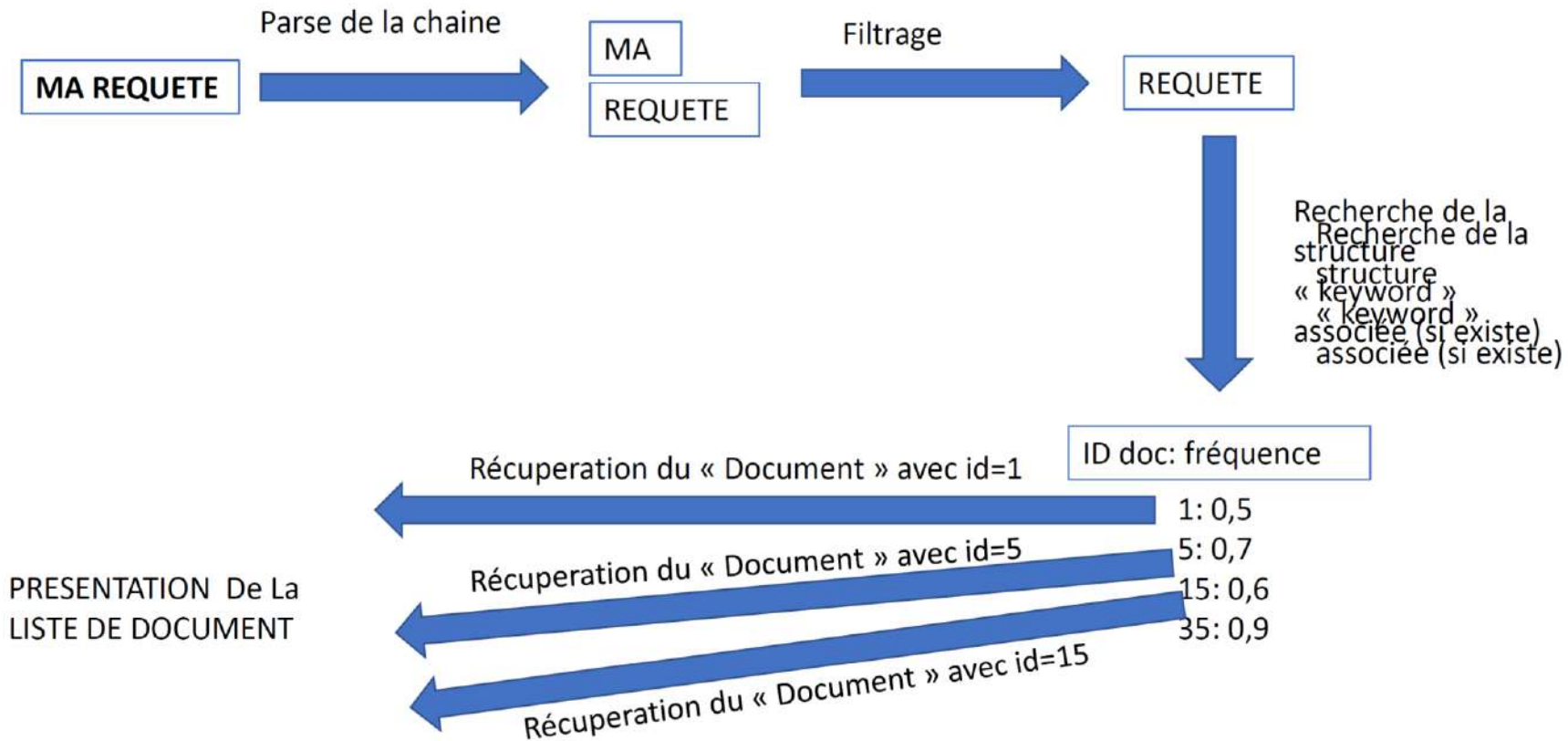
R5.C.07: *DONNEES MASSIVES*

Support TP1



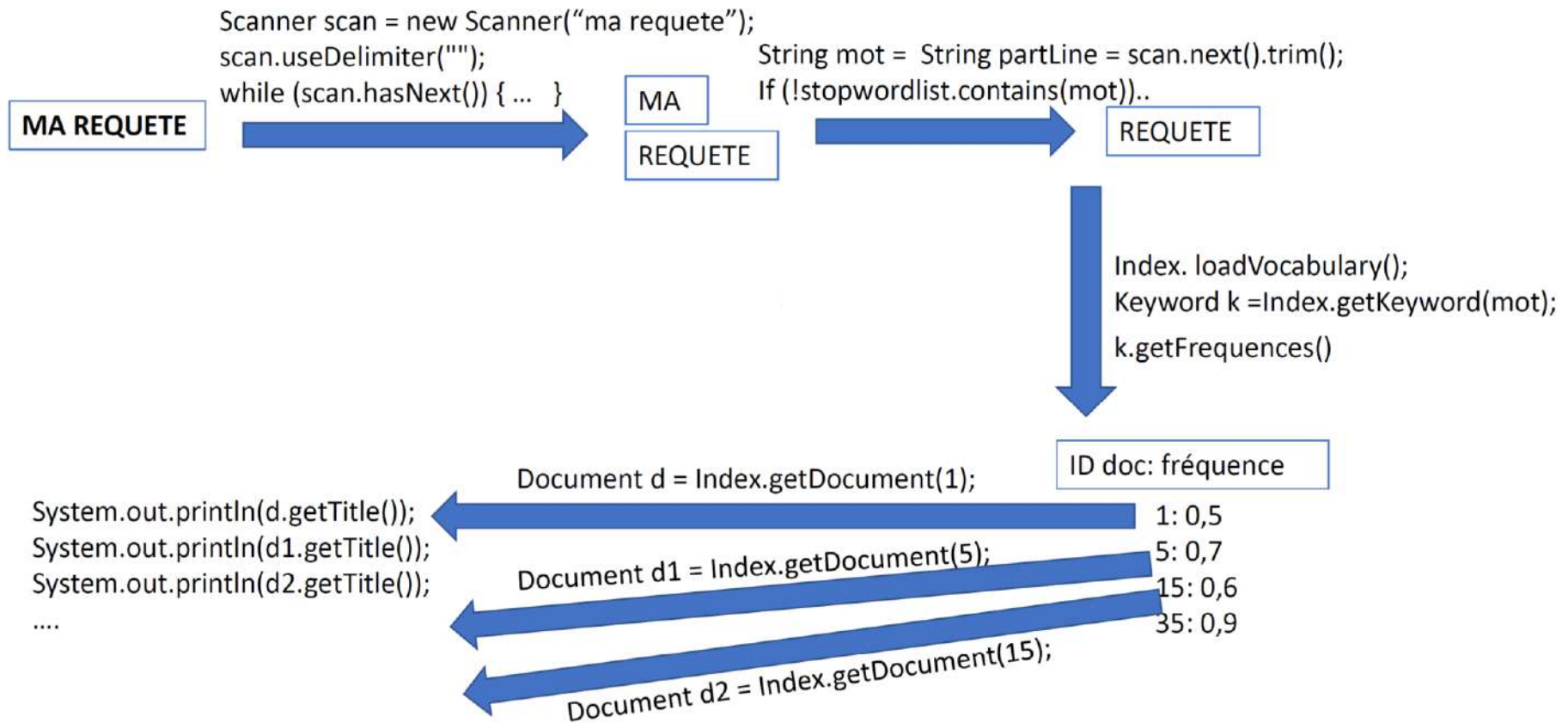
MODELE VECTORIEL

OPERATIONS A EFFECTUER



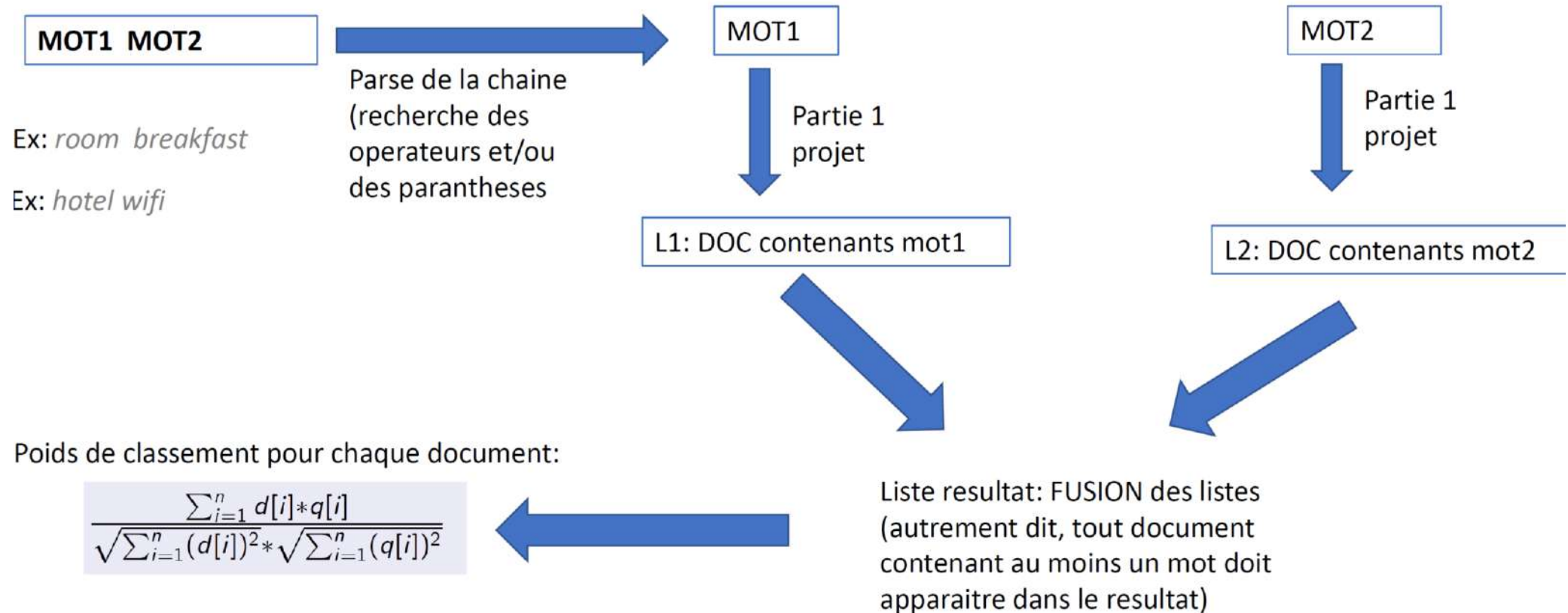
MODELE VECTORIEL

TRADUCTION en JAVA



MODELE VECTORIEL

OPERATIONS A EFFECTUER



MODELE VECTORIEL

OPERATIONS A EFFECTUER

$$\frac{\sum_{i=1}^n d[i]*q[i]}{\sqrt{\sum_{i=1}^n (d[i])^2} * \sqrt{\sum_{i=1}^n (q[i])^2}}$$

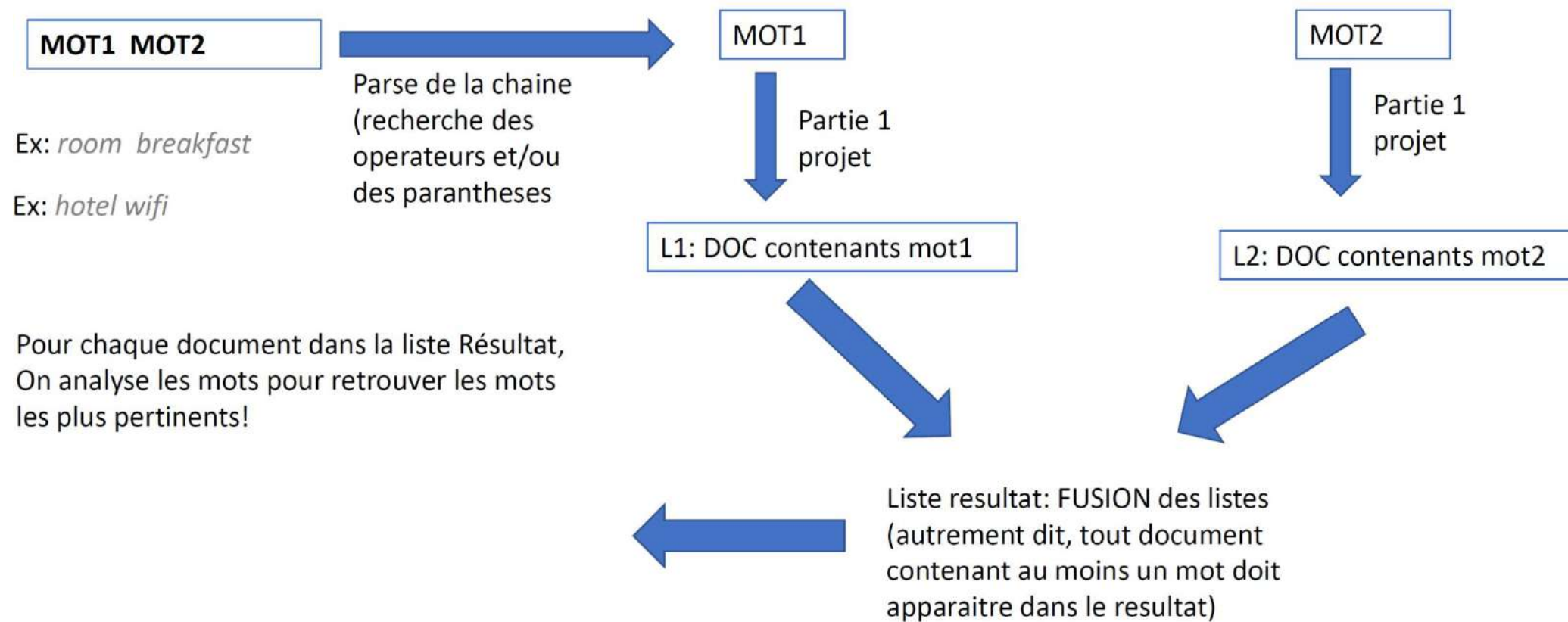
Astuce: il faut appliquer cette formule-là seulement pour les dimensions en commun! (autrement dit, pour chaque document Dans la liste résultat, on applique la formule seulement en considérant les mots en commun entre le document et la requête).

	Room	Break fast	Swim	Beautiful	Wifi	bar
Requete (room wifi bar)	0,33	0	0	0	0,33	0,33
Doc1	0,5	0.6	0.4	0.2	0.3	0

$$\text{Valeur classement doc1} = (0.5*0.33)+(0.3*0.33)/ \sqrt{[(0.5*0.5)+(0.6*0.6) \\ (0.4*0.4)+(0.2*0.2)+(0.3*0.3)]} * \sqrt{(0.33*0.33)+(0.33*0.33)+(0.33*0.33)}$$

NB: pour les autres dimensions, si un coeff est = 0, le partiel est quand même à 0 (du coup pas la peine de le prendre en compte!)

BIR



BIR

Les fréquences considérées sont modifiées de la façon suivante:

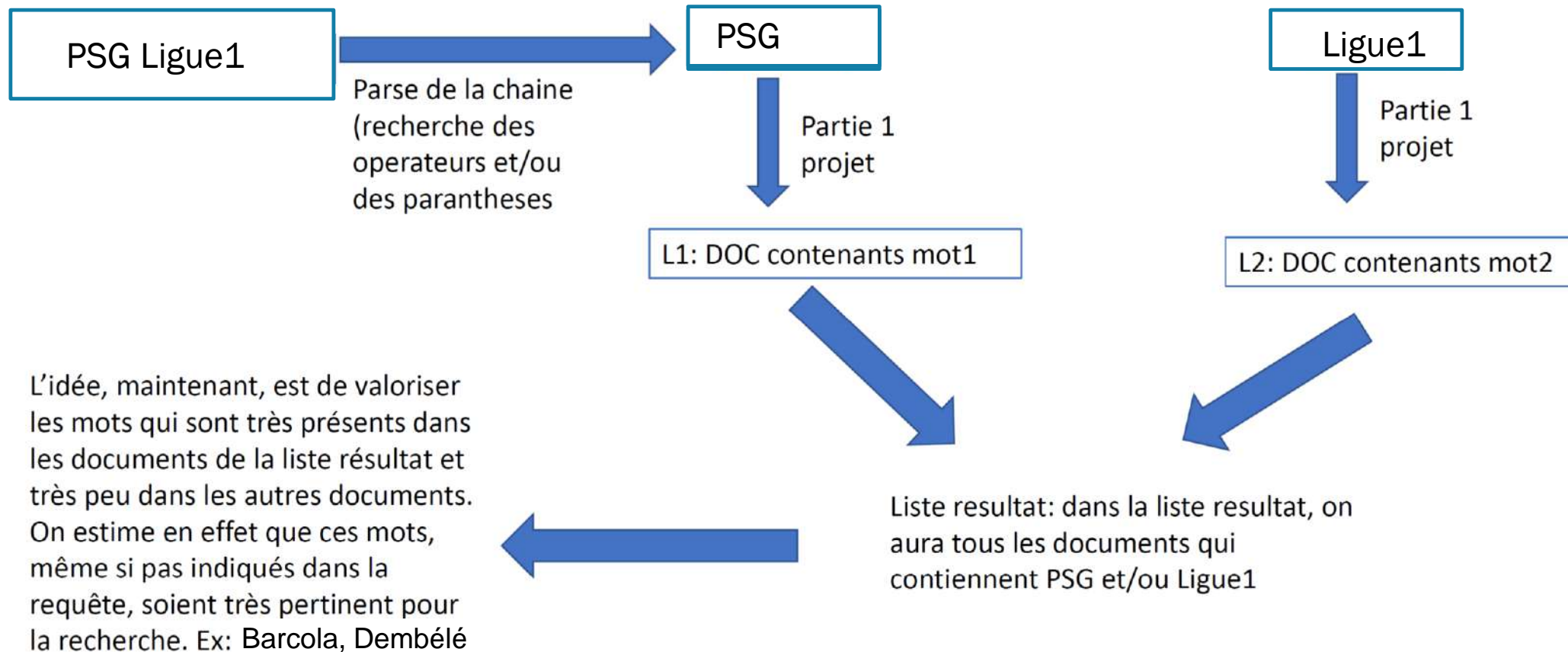
$$\text{Valeur de pertinence (relevance Feedback) =} \\ \log \left\{ \frac{r / (R - r)}{(n - r) / (N - n - R + r)} \right\} * \\ | (r - R) - (n - r) / (N - R) |$$

Ou:

- r: nombre de doc qui contiennent le mot dans liste Résultat
- n: nombre total de doc du corpus qui contiennent le mot
- R: nombre de doc in liste Résultat
- N: nombre de doc total

Les mots avec la valeur de pertinence plus élevée (avec un seuil) sont rajoutés à la requête!

BIR



BIR

