K smallest edit distances

Let $lev_{a,b}(i,j)$ be the set of k smallest edit distances between the first i characters of the string a and the first j characters of the string b.

Let's introduce the following sets:

$I_{a,b}(i,j) := \{d+1 \mid d \in lev_{a,b}(i-1,j)\}$	The set of k edit distances obtained through insertions
$D_{a,b}(i,j) := \{d+1 \mid d \in lev_{a,b}(i,j-1)\}$	The set of k edit distances obtained through deletions
$S_{a,b}(i,j) := \{d+1_{(a_i \neq b_j)} \mid d \in lev_{a,b}(i-1,j-1)\}$	The set of k edit distances obtained through substitutions
$E_{a,b}(i,j) := I_{a,b}(i,j) \cup D_{a,b}(i,j) \cup S_{a,b}(i,j)$	The set of all $3k$ edit distances

Where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$ and equal to 1 otherwise.

Let's consider the set $E_{a,b,k}(i,j) \subseteq E_{a,b}(i,j)$, such that $|E_{a,b,k}(i,j)| = k$ (in case if $|E_{a,b}(i,j)| < k$ then $E_{a,b,k}(i,j) = E_{a,b}(i,j)$), and for every $x \in E_{a,b,k}(i,j)$ and for every $y \in E_{a,b}(i,j) \setminus E_{a,b,k}(i,j)$ it follows that $x < y$.

Then the set of k smallest edit distances between the first i characters of the string a and the first j characters of the string b is defined as follows: $lev_{a,b}(i,j) := E_{a,b,k}(i,j)$.

Yurii Lahodiuk, 26.11.2017

Proof of correctness: (proof by the smallest counterexample)

For the sake of convenience let's introduce an auxiliary notation: $min_k(X)$ for some set X with natural numbers, and for some $k \in \mathbb{N}$, which means that $min_k(X) \subseteq X$ and $\forall a \in min_k(X), \forall b \in X \setminus min_k(X)$ it follows, that $a < b$, and $|min_k(X)| = k$ (in case if $|X| < k$, then $min_k(X) = X$).

Induction Basis:
Induction Hypothesis:

- The set $lev_{a,b}(i-1,j)$ contains the k smallest edit distances between the first $i-1$ characters of the string a and the first j characters of the string b.
- The set $lev_{a,b}(i,j-1)$ contains the k smallest edit distances between the first i characters of the string a and the first $j-1$ characters of the string b.
- The set $lev_{a,b}(i-1,j-1)$ contains the k smallest edit distances between the first $i-1$ characters of the string a and the first $j-1$ characters of the string b.

Inductive Step:

We want to show, that the *Induction Hypothesis* implies, that the set $lev_{a,b}(i,j)$ contains the k smallest edit distances between the first i characters of the string a and the first j characters of the string.
For the sake of contradictions let's assume, that there there exists an edit distance $y \notin lev_{a,b}(i,j)$ between the first i characters of the string a and the first j characters of the string, such that $\exists x \in lev_{a,b}(i,j)$ for which $y < x$ (thus the set $lev_{a,b}(i,j)$ doesn't contain the k smallest edit distances):

$$\exists y \notin lev_{a,b}(i,j) \land \exists x \in lev_{a,b}(i,j) : y < x$$

Let's expand the expression in a following way:

$$\exists y \notin lev_{a,b}(i,j) \land \exists x \in lev_{a,b}(i,j) : y < x \Leftrightarrow \quad \text{By definition of } lev_{a,b}(i,j)$$

$$\Leftrightarrow \exists y \notin min_k\Big(I_{a,b}(i,j) \cup D_{a,b}(i,j) \cup S_{a,b}(i,j)\Big) \land \exists x \in lev_{a,b}(i,j) : y < x \Rightarrow \quad \text{As far as } y < x$$

$$\Rightarrow \exists y \notin I_{a,b}(i,j) \cup D_{a,b}(i,j) \cup S_{a,b}(i,j) \land \exists x \in lev_{a,b}(i,j) : y < x \Leftrightarrow \quad \text{By definition of } I_{a,b}(i,j),\ D_{a,b}(i,j),\ S_{a,b}(i,j)$$

$$\Leftrightarrow \exists y : \Big((y-1) \notin lev_{a,b}(i-1,j)\Big) \land \Big((y-1) \notin lev_{a,b}(i,j-1)\Big) \land$$

$$\land \Big((y - 1_{(a_i \neq b_j)}) \notin lev_{a,b}(i-1,j-1)\Big) \land \exists x \in lev_{a,b}(i,j) : y < x$$

According to the *Induction Hypothesis* - the set $lev_{a,b}(i-1,j)$ is the set of the k **smallest edit distances**, thus from $(y-1) \notin lev_{a,b}(i-1,j)$ it follows, that $\forall a \in lev_{a,b}(i-1,j) \Rightarrow (y-1) > a$ (and the same logic is applicable to the sets $lev_{a,b}(i,j-1)$ and $lev_{a,b}(i-1,j-1)$):

$$\exists y : \Big((y-1) \notin lev_{a,b}(i-1,j)\Big) \land \Big((y-1) \notin lev_{a,b}(i,j-1)\Big) \land$$

$$\land \Big((y - 1_{(a_i \neq b_j)}) \notin lev_{a,b}(i-1,j-1)\Big) \land \exists x \in lev_{a,b}(i,j) : y < x \Rightarrow$$

$$\Rightarrow \exists y : \Big(\forall a \in lev_{a,b}(i-1,j) \Rightarrow y > a+1\Big) \land \Big(\forall b \in lev_{a,b}(i,j-1) \Rightarrow y > b+1\Big) \land$$

$$\land \Big(\forall c \in lev_{a,b}(i-1,j-1) \Rightarrow y > c + 1_{(a_i \neq b_j)}\Big) \land \exists x \in lev_{a,b}(i,j) : y < x \Rightarrow$$

$$\Rightarrow \exists y : \Big(\forall x \in lev_{a,b}(i,j) \Rightarrow y > x\Big) \land \exists x \in lev_{a,b}(i,j) : y < x \Rightarrow$$

$$\Rightarrow Contradiction.$$

■