مسئله در مورد تصحیح اشتباهات چاپی در متن بدون استفاده از دیکشنری است. به شما متن حاوی تعداد زیادی خطای چاپی داده می شود و هدف این است که تا حد امکان بسیاری از خطاها را اصلاح کنید.

در این مسئله وضعیت ها به حروف درستی که باید تایپ شود اشاره دارد و خروجی به حروف واقعی که تایپ شده اشاره دارد.

دنباله ای از خروجی ها داده شده(به عنوان مثال حروف تایپ شده واقعی) و مسئله ساخت دوباره دنباله حات های مخفی است

(دنباله ای از حروف مورد نظر) .بنابراین داده ها برای مسئله زیر به شکل زیر خواهد بود:

```
i i
n n
t t
r r
o o
d x
u u
c c
t t
i i
o i
n n
_ _
t t
h h
e e
```

که ستون چپ حروف درست و ستون سمت راست شامل متن با خطا است.

داده ها برای این مسدله به صورت زیر ایجاد شده.با یک فایل متنی شروع می کنیم.برای سادگی تمام اعداد و نشانه گزاری ها به فاصله سفید و تمام حروف به حالت حروف کوچک تبدیل شده است.متن باقی مانده دنباله ای از حروف کوچک و کارکتر فاصله است که با علامت – در فایل داده مشخص شده است.در ادامه علط های چاپی با صورت مسنوعی به صورت زیر به داده ها اضافه شده است:

با احتمال 90 درصد حرف نوشته می شود اما با احتمال 10 درصد ،یک همسایه تصادفی(در در ترتیب فیزیکی صفحه کلید)از حرف به جای حرف درست نوشته می شود.کارکتر فاصله همیشه به صورت درست نوشته می شود.در یک نوع سخت تر از مسئله نرخ خطا نرخ خطا به 20 درصد افزایش می یابد.

تقریبا 20000 کاراکتر اول فایل برای تست کنار گذاشته شده است.161000 کارکتر بقیه برای آموزش استفاده می شود.

به عوان مثال،فایل اصلی شروع می شود با :

```
introduction the industrial revolution and its consequences have been a
disaster for the human race they have greatly increased the life expectancy
of those of us who live in advanced countries but they have destabilized
society have made life unfulfilling have subjected human beings to
indignities have led to widespread psychological suffering in the third world
to physical suffering as well and have inflicted severe damage on the natural
```

world the continued development of technology will worsen the situation it will certainly subject human beings to greater indignities and inflict greater damage on the natural world it will probably lead to greater social disruption and psychological suffering and it may lead to increased physical suffering even in advanced countries the industrial technological system may survive or it may break down if it survives it may eventually achieve a low level of physical and psychological suffering but only after passing through a long and very painful period of adjustment and only at the cost of permanently reducing human beings and many other living organisms to engineered products and mere cogs in the social machine

با خطای 20 درصد،شبیه به متن پایین خواهد بود:

introductipn the industfial revolhtjon and its consequences bafw newn a diszster rkr the yumab race thdy have grwatky increased the ljte esoectandy od thosr of is who libe in advanced coubfries but they have fewtabipuzee xociwty have made life ujfuorillkng have wubjwdted humah beints to incihbjtids have led to qidespreze lsyxhllotical shffeding kn tne third wkrld to phyxicql sufcefimg as weol and hqve ingoidtex srvere damsge on the natural world the confinued developmeng of twvhjllogy will wotsen thd situation it wull certaknly sunjrct yyman beingw tl greater ibdignities snd infpixt greagwr damsge on fhe natural alrld it wjlk probably lwad tk grezter sofiqp disrupgiln and pstchokofucal wufterkng anc it may kead fl uncreqxed pgusiczl sucfreinh even in acgajved countries the indhsteial tedhnologicak system may survivr or ut nay brezk down uf it survives it nay evenyuakly achieve a los lwvel of phyxkcal and psycyoligical sufveribg but only after passing theough a long amd very painful periox od adjuwtmebt and only at the fost kf permsnently reducing hymaj veings abs nsjy otgwr kuving orbanisms to envineered leoduxfs amd mere clgs in thr soxiap maxhjne

نرخ خطلا حدود 16.5 درصد است(کمتر از 20 درصد چون کاراکتر های فاصله درست تایپ شده است.)

متن دوباره ساخته شده به وسیله HMM با الگوریتم ویتربی شبیه یه متن زیر است:

introduction the industrial revoluthon and its consequences bare neen a dissster ror the tuman race they have greatly increased the lite esoectandy od those of is who libe in advanced counfries but they have festabupusee cocisty have made live intiorilling have wibjested human beints to incingitids have led to widesprese lsysullotical suffeding in the third world to physical surcefing as weol and have ingoistes severe damage on the natural world the continued developmeng of techillogy will wotsen the situation it will certaknly sunirct tyman beinge tl greater indithities and infoist greager damage on the natural aleld it will probably owad to grester sofial distuption and pstchomofucal wiftering and it may kead fl increqxed ogusical suctreing even in achanved countries the industeial technologicak system may survive or ut nay break down if it survives it nay eventually achieve a los level of physical and psycholigical survering but only arter passing theough a long and very paindul perios od adjustment and only at the fost of permanently reducing human veings ans nany other kiving organisms to envineered leodusts and mere clys in the social machine

نرخ خطا به حدود 10.4 درصد کاهش یافت.

داده های مربوط به مقدار دهی در typos10.data و ,typos20.data که شامل داده ها با خطای به ترتیب 10 درصد و 20 درصد است.