



XXI АСР – ИНТЕЛЛЕКТУАЛ АВЛОД АСРИ

*Ёш олимлар ва
талабаларнинг
республика илмий-
амалий анжумани
материаллари*

Тошкент
2015 йил 19-20 ноябрь

1606. ИЖТИМОЙ-ГУМАНИТАР ФАНЛАР

МАТНЛАРНИ АВТОМАТИК ТАҲРИР ВА ТАҲЛИЛ ҚИЛИШ ДАСТУРИНИНГ ЛИНГВИСТИК БОСҚИЧЛАРИ

М. Абжалова,
НавДКИ

Ўзбек тилидаги матнларни автоматик равишда таҳрир ва таҳлил қилиш дастурини яратиш бугунги куннинг долзарб масаласига айланган. Бундай дастур тизимининг катта имконият ва юқори сифати лингвистик модулларнинг мукамал ишлаб чиқилгани билан белгиланади, улар ўз навбатида дастурнинг босқичларини ташкил этади.

Лингвистик таҳрир – бу турли кўриниш (илмий, бадий, публицистик ва расмий услублар)даги матнларнинг орфографик, грамматик, стилистик ҳамда мантиқий қурилишдаги хатоларини тўғрилаш дегани.

Матнларни таҳрир қилиш икки усулда амалга оширилади:

1. Инсон томонидан амалга ошириладиган таҳрир (бевосита таҳрир – муҳаррирлик иши).

2. Замонавий техника воситасида амалга ошириладиган таҳрир (билвосита, яъни автоматик таҳрир).

Кейинги пайтларда ахборот алмашинувининг жадаллашиши ва маълумотлар кўламининг табора кенгайиб кетаётгани сабабли билвосита таҳрирға бўлган талаб ошиб бормокда.

Автоматик таҳрир компьютер лингвистикасининг асосий йўналишларидан бири бўлиб, у компьютердаги матн муҳаррирлари (Microsoft Word, Exsel, Wordpad, Lexicon) ривожланиши билан боғлиқ ҳолда юзага келган. *Автоматик таҳрир* деганда, компьютерға киритилаётган матннинг механик хатолари автоматик тарзда тўғриланиши ва хато эканлиги ҳақидаги сигналларнинг фойдаланувчига таклиф этилиши назарда тутилади. Масалан, Microsoft Word дастури асосида инглиз тили ёки рус тилида матн ёзсак, унда сўзларнинг ёзилиши билан боғлиқ хато-

лар бўлса, компьютернинг ўзи уларнинг остига қизил ёки яшил рангдаги тўлкинли чизик чизади; катта ҳарф билан ёзилиши керак бўлган сўзларни кичик ҳарф билан ёзсак, уларни автоматик равишда бош ҳарфга ўтказади; бандларни автоматик равишда рақамлаб беради; хатбошини ўзи қўяди ва ҳоказо. Бу эса ишни осонлаштиради.

Бирок ҳозирги пайтга қадар мукаммал автоматик таҳрир қиладиган лингвистик дастур яратилмаган. Компьютер таҳрир дастури таҳрирга қисман ёрдам беради, холос. Аниқроқ қилиб айтганда, матнни синтактик таҳрир қилиш муаммоси ўз ечимини топмаган. Синтактик таҳрирнинг мураккаблиги шундаки, матнда кўчма маъноли сўзлар, иборалар, тагмаъноларга эга ишлатилиши мумкин.

Матнларни автоматик таҳрир қилишнинг мукаммал ва юқори сифатли тизимига эришиш учун унга киритилаётган матнларни нафақат орфографик ва морфологик жиҳатдан, балки синтактик (сўз бирикмаси ва гап таҳрири), семантик (матнда қўлланилаётган тушунчалар таҳрири) ва прагматик (тушунчаларнинг тўғри қўлланиши ва уларни қўллаш мақсади таҳрири) таҳрир қилиш имкониятларини яратиш лозим.

Матнларни тўлиқ автоматик таҳрир ва таҳлил қилиш учун дастур қуйидаги модулларга эга лингвистик босқичлардан таркиб топиши лозим:

Графематик таҳрир босқичида *матндаги хатбоши, рақам, пунктуацион* ва бошқа махсус рамзий белгилар аниқланади, яъни сўз ва сўзшаклларга қадар текширувни амалга оширади. Графематик таҳрир босқичининг мақсади – матндаги энг кичик бирликларни ҳам аниқлаш ва таснифлашдан иборат. Бундай бирликларга қуйидагилар қиради: сўз, хатбоши, тиниш белгилари, саналар, пул бирликларининг рамзлари, сўз-рақамли бирикмалар, сонлар, IP-манзиллар ва файл номлари, телефон рақамлари. Графематик таҳрир кўп босқичли таҳлил қилиш дастурининг илк таҳрир босқичи ҳисобланиб, у кейинги босқичларнинг самарали ва тўлақонли бўлиши учун асос яратди.

Морфологик таҳрир босқичи асос ва шу асосдан юзага келган сўзшаклларни текширади. Морфологик таҳрирнинг мақсади – фақат орфографик хатоларни текширишдан иборат, деб ҳисоблаш унинг вазифа кўламини торайтиради. Ушбу босқичда киритилаётган сўзшаклнинг қандай асосий шакл-

дан ва қайси лингвистик параметрлардан юзага келганини аниқлайди. Бу кейинчалик бир сўзнинг барча сўзшакллари устида ишлашга эмас, балки асосий шаклнинг қандай лингвистик параметрларни олиш имкониятига эгаллиги устида иш олиб боришга туртки беради. Масалан, от туркумидаги *китоб* сўзининг қандай шакл ҳосил қилувчи ва синтактик шакл ясовчи қўшимчаларни қабул қилиши ҳамда уларнинг комбинацияларини бириктириб, сўзшаклларини ҳосил қилиш имконияти, яъни параметрлари дастурнинг базасига киритилади. Натижада *китоб* сўзининг барча шакллари киритилиб, дастур таъминотидан катта ҳажм эгалланмайди, балки асосий шакл ва унга бирикиши меъёрий ҳисобланган лингвистик параметрлар киритилиб, ҳажм иқтисод қилинади.

Синтактик таҳлил босқичи – матн таҳрирининг энг қийин босқичи. Унинг асосий вазифаси сўзларнинг ўрнини ва уларнинг ўзаро бирикишини аниқлаш ҳисобланади¹. Шу боис ушбу босқични *таҳрир босқичи* эмас, балки лингвистик *таҳлил босқичи*, дейиш ўринли. Бир сўзшакл бир қанча сўзнинг грамматик шаклига тўғри келиши мумкин. Бундай ҳолда сўзларни фарқлаш учун матн киритувчига шундай сўзлар варианты кўрсатилади ва матн терувчи сўзлардан мувофиғини танлаб қўяди. Ушбу жараён морфологик таҳрирда ҳам кузатилади. Яъни, сўзнинг хато ёзилгани аниқлангач, матн терувчига танлаб қўйиш имконини берувчи сўзлар вариантини таклиф этади. Синтактик таҳлилда асосий мақсад сўзларнинг ўзаро боғланиш меъёрларини эътиборга олишдир.

Матнларни грамматик таҳрир қилишда (Grammar checking) компьютерга киритилаётган матндаги гап структурасининг грамматик жиҳатдан тўғри тузилганлиги текширилади. Бунинг учун дастурга гапларнинг комбинацияларини киритиш талаб этилади. Шунда гапни дастур киритилган кўплаб меъёрий комбинациялар доирасида текширади. Агар гап тузилиши комбинацияларга мувофиқ келмаса, дастур уни хато деб ҳисоблайди ва фойдаланувчига бу ҳақда хабар беради. Grammar checking дастурида фақат грамматик меъёрлар эмас, балки семантика, сўзларнинг маъноси ҳамда гапнинг мантикийлигига ҳам эътибор берилади. Масалан, *Қуш учмоқда. Одам учмоқда. Девор*

¹ *Большакова и др.* Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. – М.: 2011. с.106.

учмоқда гаплари грамматик жиҳатдан тўғри ёзилган, яъни эга + кесим муносабати мавжуд (III шахс, бирлик). Аммо семантикасига эътибор берсак, гаплар хато тузилган. Чунки одамнинг қаноти йўқ, фақат бадиий матндагина инсоннинг учиши ҳақида ёзиш мумкин. Матнларни компьютер воситасида қайта ишлашда илмий, расмий-идоравий, ахборот доирасидаги матнлар ҳам эътиборга олинади¹. Келтирилган *Девор* учмоқда гапининг семантикаси эса мутлақо нотўғри ва гап ҳеч қайси услубга ҳос эмас. Чунки девор муқим бир жойда турувчи, бинонинг, уйнинг томини кўтариб турувчи ёки хоналарга ажратувчи тик кўтарилган қисми².

Матнларни семантик таҳлил қилиш босқичида тушунчага эга сўзнинг ўринли қўлланилаётгани текшириб борилади. Бунинг учун дастур таъминотига тезаурус луғатларни киритиш талаб қилинади. Тезаурус (юнонча *thesaurós* – хазина, бойлик) луғатнинг ўзига ҳос махсус кўриниши бўлиб, унда лугавий бирликлар ўртасидаги семантик муносабат (синонимлар, антонимлар, паронимлар, гипонимлар, гиперонимлар ва бошқалар) кўрсатилган бўлади³.

Замонавий технологияларнинг кундалик турмушимизда муқим ўрин эгаллаши ва илм-фаннинг ривожланиши автоматик таҳрир ва таҳлилнинг аҳамиятини оширмақда. Автоматик таҳлил фақат матнлардаги механик хатоларни бартараф этибгина қолмай, матннинг саводли ёзилишини ҳам таъминлайди, матн киритувчига жумлада хато борлиги ҳақида хабар бериб, уни хатони бартараф этишга ундайди, натижада тўлиқ лингвистик меъёрлар даражасидаги матнни юзага келтиришда вақт тежалади.

Бундай имкониятга эришиш учун, албатта, ўзбек тили асосида кўп босқичли автоматик таҳлил дастурининг лингвистик таъминотини мукамал даражада ишлаб чиқиш талаб этилади.

¹ Абжалова М.А. Ўзбек тилидаги матнларни автоматик таҳрирлашда услуб масаласи // Таълим тизимида ижтимоий-гуманитар фанлар. № 2, Т.: 2013, 141–145-бетлар.

² Қ а р а н г : Ўзбек тилининг изоҳли луғати. 1-жилд. Т., 2006., 588-б.

³ Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы – М., 2006. с. 149–158.